Le patrimoine nativement numérique : enjeux de collecte et de recherche

EGIDE 2018
14 mars 2018, Lille

Sara AUBRY (Bibliothèque nationale de France)
Valérie SCHAFER (C²DH, Université du Luxembourg)

BnF Bibliothèque nationale de France

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Le web est éphémère.

Les archives du web servent à retrouver…

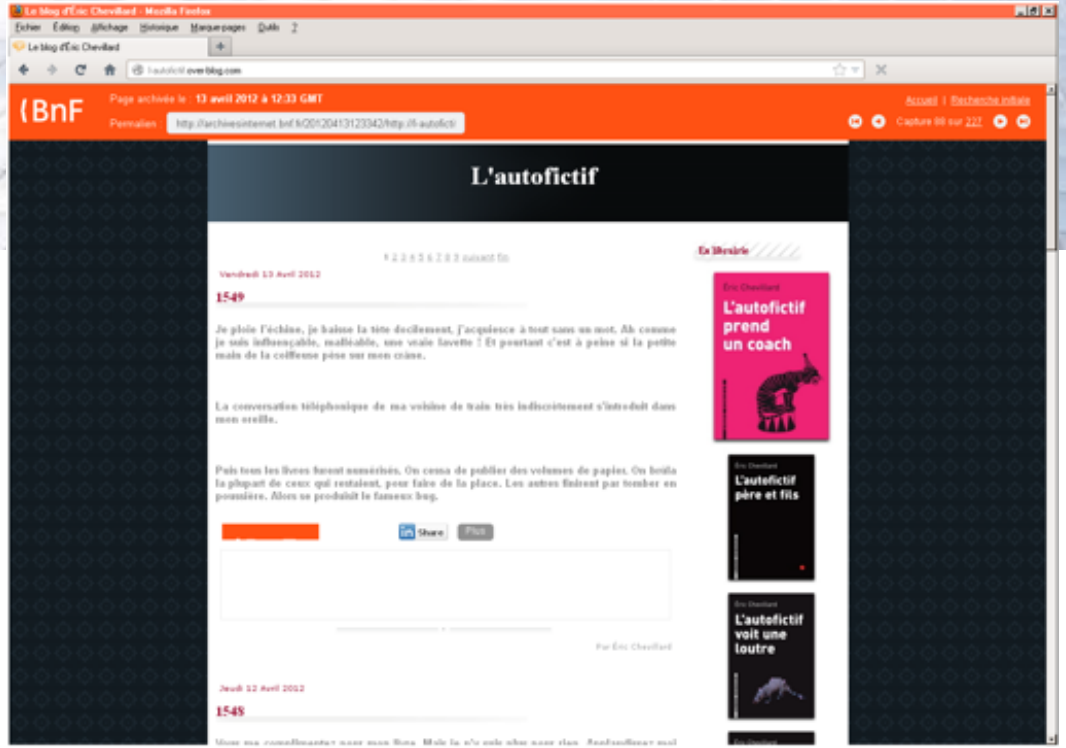
BnF

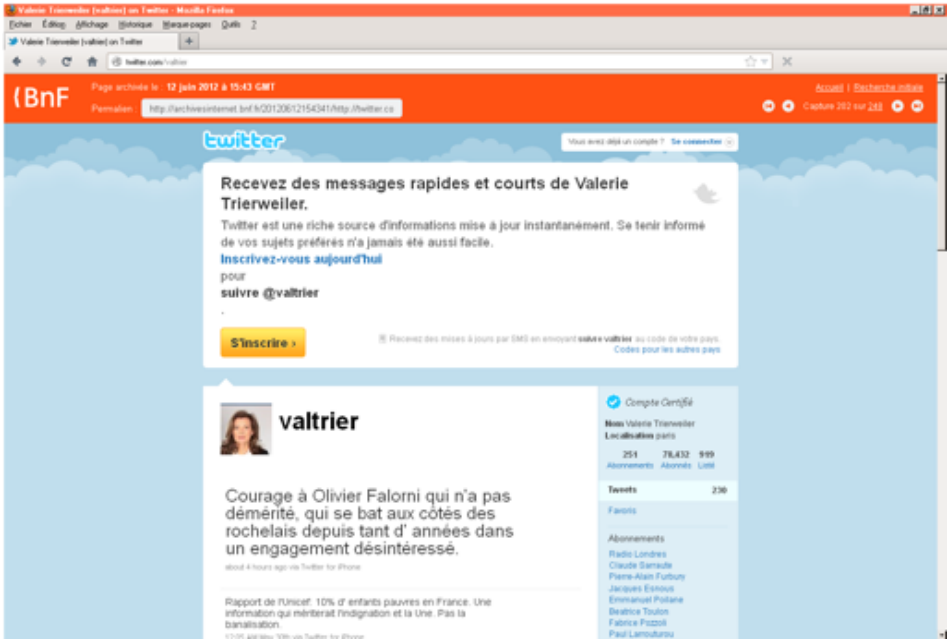
BnF
Page archivée le : 30 octobre 2014 à 13:20 GMT
Permalien : http://archivesinternet.bnf.fr/20141030132025/http://www.fiac
Accueil | Recherche initiale
Capture 13 sur 62
GRAND PALAIS
PARIS
23-26 OCTOBRE 2014
fiac!
CONVENTION CENTER
LOS ANGELES
27-29 MARCH 2015

BnF
Page archivée le : 10 juin 2016 à 08:02 GMT
L'ÉQUIPE
PREMIUM
EURO 2016
FOOTBALL
TENNIS
RUGBY
SPORTS
EXPLORE
#NTF
VIDÉO
AGENDA TV
Directs
Résultats
EURO LE GUIDE
EURO MEMORIES
TENNIS MATCHES
EXPLORE ISLANDE
VOS PRONOS POUR L'EURO
JO DE RIO 2016
«JE N'AIME PAS ÊTRE LE NUMÉRO 2»
FOOT
ÉQUIPE DE FRANCE
FRANCE
21h00
ROUMANIE
CHRONO
VOIR TOUT LE CHRONO
LE QUOTIDIEN

BnF
Page archivée le : 10 janvier 2015 à 16:05 GMT
Permalien : http://archivesinternet.bnf.fr/20150110160546/http://www.thea
Accueil | Recherche initiale
Capture 28 sur 47
JE SUIS CHARLIE
CHIALER
CHARLIE CHERCHE SA PROCHAINE UNE
ENTREZ
S'inscrire à la newsletter ICI

BnF

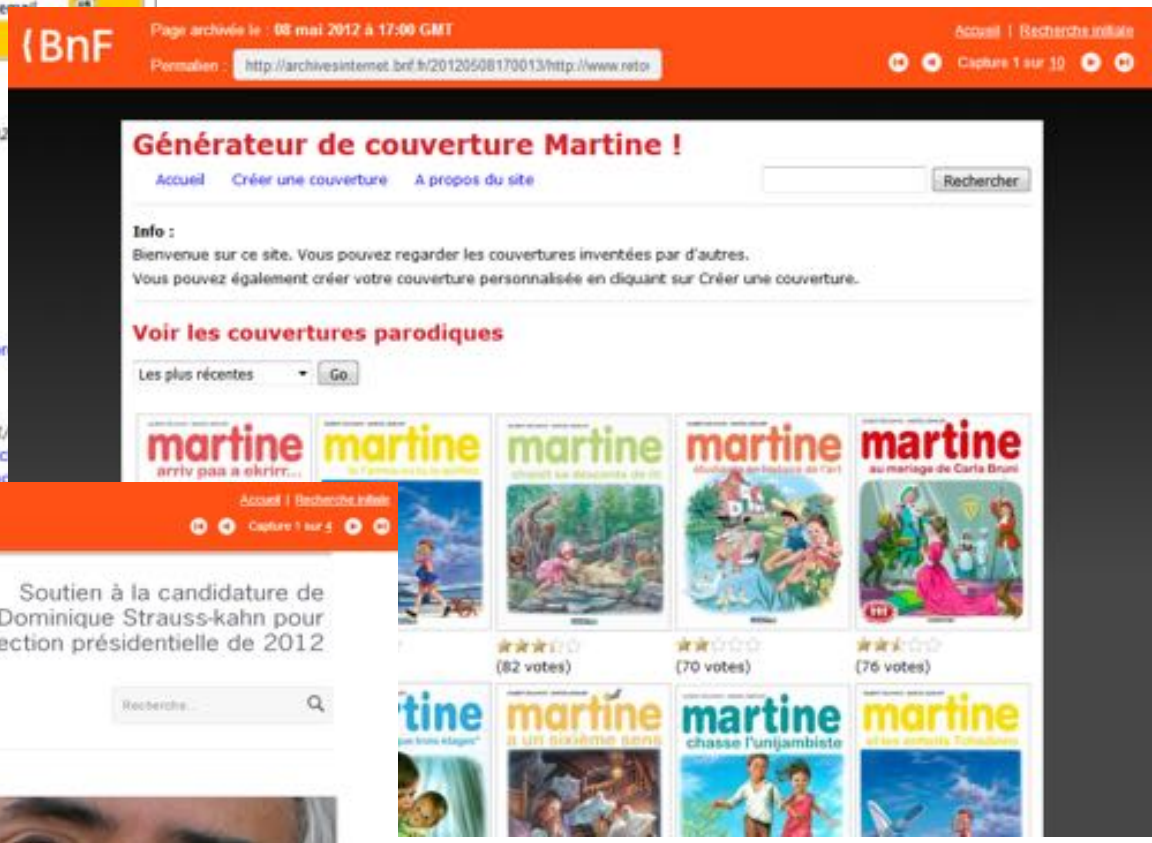

(BnF
L'autofictif
1549
L'autofictif prend un coach
L'autofictif père et fils
L'autofictif voit une loutre
1548


(BnF
twitter
Recevez des messages rapides et courts de Valerie Trierweiler.
Twitter est une riche source d'informations mise à jour instantanément. Se tenir informé de vos sujets préférés n'a jamais été aussi facile.
Inscrivez-vous aujourd'hui
pour
suivre @valtrier
S'inscrire ›
valtrier
Courage à Olivier Falorni qui n'a pas démérité, qui se bat aux côtés des rochelais depuis tant d' années dans un engagement désintéressé.

(BnF
mois offerts !
Découvrez l'offre !
MEDIAPART
CENSURÉ
Bettencourt : l'affaire que vous ne pouvez plus lire
Une justice injuste
Maistre censeur sur son lignage perché
La liste des articles censurés
La jurisprudence constante de la Cour européenne

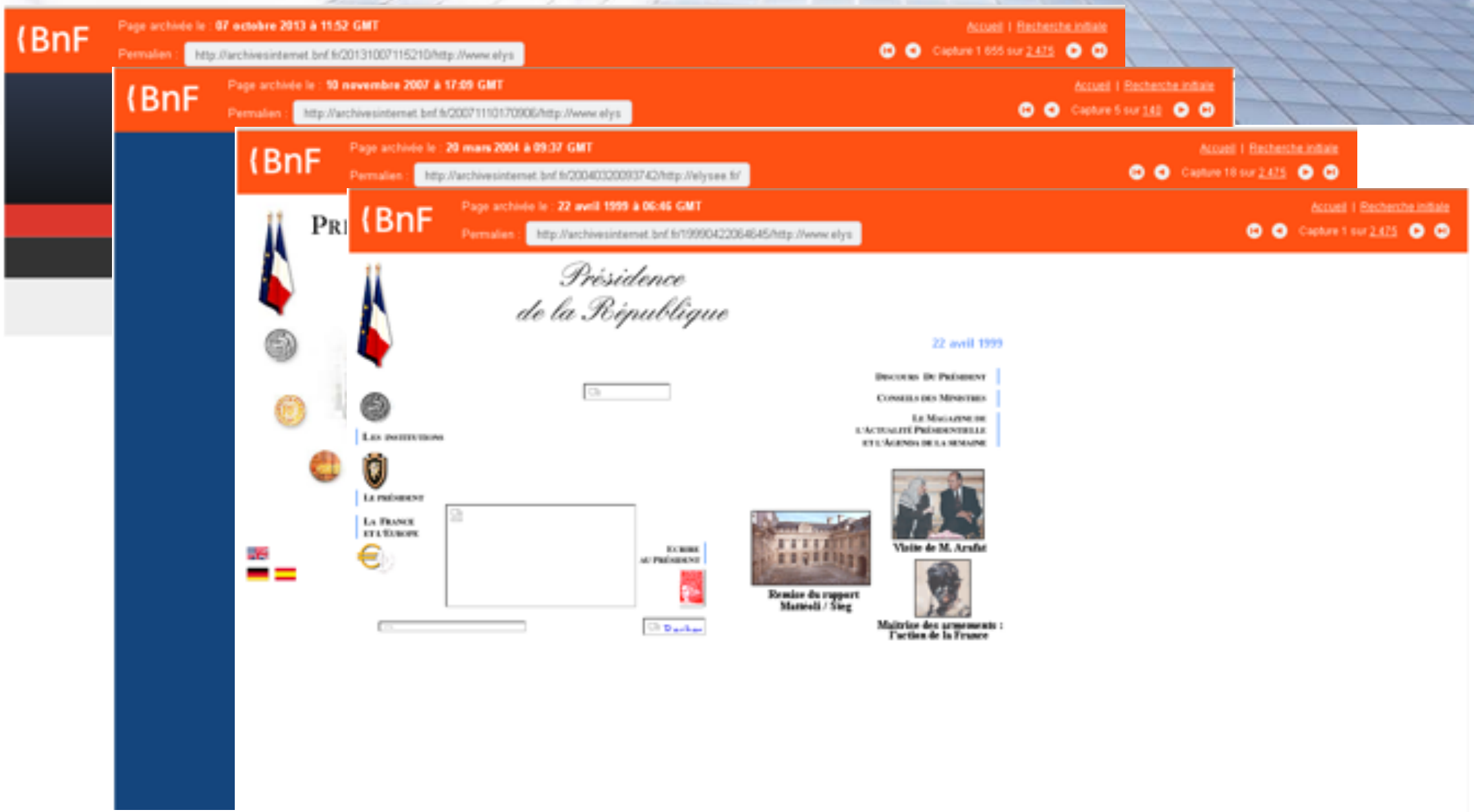
(BnF

BnF
Page archivée le : 30 mars 2002 à 11:55 GMT
Permalien : http://archivesinternet.bnf.fr/20020330115516/http://mauvaisg
Accueil | Recherche initiale
Capture 5 sur 58
Mauvais genres.bib
mise à jour le 29 mars 2002
Rechercher un article
OK
Toutes les rubriques
REVUE DE PRESSE
AGENDA
LISTE DE DISCUSSION
Forte de 640 abonnés, cette liste est devenue une source d'information incontournable sur le polar et la SF : critiques, liens, annonces, débats, rencontres avec des auteurs.
Pour en savoir plus
Pour s'inscrire
DERNIÈRES MISES EN LIGNE
Critiques
Sandrone et associes. Sandrone se soigne / Sandrone Dazieri. - Métailié, 2001 et 2002 29/03/02
1974/ David Peace. - Rivages, 2002 29/03/02
Studio Sex / Liza Marklund. - Livre de poche, 2002 28/03/02
Requiem pour Philip K. Dick / Michael Boshop. - Gallimard, 2002 28/03/02
Le Fantastique dans tous ses états / Roger Bozzetto. - Publications de l'Université de Provence, 2001 25/03/02
Mystic river / Dennis Lehane. - Rivages, 2002 25/03/02
La Fille aux cheveux noirs / Philippe K Dick. - Gallimard, 2002 21/03/02
Malhorne, épisode 1 : Le Trait d'union des mondes / Jérôme Camut. - Serpent à plumes, 2002 20/03/02
La Surprise du chef / Anthony Bourdain. - C. Bourgois, 2002 20/03/02
L'Artiste des Dames / Eduardo Mendoza. - Seuil, 2002 19/03/02
Bibliographies
Rencontres
avec Joëlle Wintrebert en avril 02
avec Yasmina Khadra en avril 02
avec Roger Martin en mars 02
avec Ayerdhal en mars 02
avec Claude Amoz en fev. 02
Interviews
Eduardo Mendoza à la librairie à Ombr
25/03/02
Portraits et commentaires
Claude Mesplède : parole de polar
Critiques
Bibliographies
Rencontres
Interviews
Portraits
Collections
Espace Jeunesse
Radio
Répertoire
Carnet
Association 813
La Noiraude
Hard-boiled Dicks
Revues, Librairies...


BnF
Page archivée le : 08 mai 2012 à 17:00 GMT
Permalien : http://archivesinternet.bnf.fr/20120508170013/http://www.reto
Accueil | Recherche initiale
Capture 1 sur 10
Générateur de couverture Martine !
Accueil
Créer une couverture
A propos du site
Rechercher
Info :
Bienvenue sur ce site. Vous pouvez regarder les couvertures inventées par d'autres.
Vous pouvez également créer votre couverture personnalisée en cliquant sur Créer une couverture.
Voir les couvertures parodiques
Les plus récentes
Go
martine
(82 votes)
(70 votes)
(76 votes)
chasse l'unijambiste

BnF
Page archivée le : 19 mai 2011 à 13:56 GMT
Permalien : http://archivesinternet.bnf.fr/20110519135621/http://www.d
Accueil | Recherche initiale
DSK 2012-bouchesdurhone.fr
Soutien à la candidature de Dominique Strauss-kahn pour l'élection présidentielle de 2012
ACCUEIL
CONTACT
LECTURES
MÉDIAS
SIGNEZ L'APPEL
Recherche...
Blog de soutien à la candidature de Dominique Strauss-Kahn pour les élections présidentielles de 2012.
Appel à la candidature de Dominique Strauss-Kahn Signez l'appel en cliquant ici.

BnF


BnF
Page archivée le : 07 octobre 2013 à 11:52 GMT
Permalien :
Accueil | Recherche initiale
Capture 1 655 sur 2 475
BnF
Page archivée le : 10 novembre 2007 à 17:09 GMT
Permalien :
Accueil | Recherche initiale
Capture 5 sur 141
BnF
Page archivée le : 20 mars 2004 à 09:37 GMT
Permalien :
Accueil | Recherche initiale
Capture 18 sur 2 475
BnF
Page archivée le : 22 avril 1999 à 06:46 GMT
Permalien :
Accueil | Recherche initiale
Capture 1 sur 2 475
Présidence
de la République
22 avril 1999
Visite de M. Arafat
Remise du rapport
Mamoli / Sieg
Maîtrise des armements :
l'action de la France

BnF

BnF Page archivée le : 21 mars 2015 à 00:35 GMT
Permalien : http://archivesinternet.bnf.fr/20150321003509/http://www.univ
Accueil | Recherche initiale
Capture 358 sur 460

BnF Page archivée le : 01 juin 2010 à 09:13 GMT
Permalien : http://archivesinternet.bnf.fr/20100601091357/http://www.univ
Accueil | Recherche initiale
Capture 41 sur 136

Accès déficients visuels | Aide en ligne | Nous contacter | Annuaire Lille 3 | Rechercher | Intranet - ENT

FUTURS ÉTU
ETUDIANTS
ANCIENS ÉT
EN REPRISE
FORMATION
PRESSE
PARTENAIRE

Accès direc
Services de l'u
UFR, Instituts
Laboratoires d
Rechercher
Inscriptions
Offre de forma
Validation des a
Publications
Maison Europée Sciences de l'Ho Société
Presses Univers Septentrion
Atelier National

BnF Page archivée le : 29 mars 2003 à 20:22 GMT
Permalien : http://archivesinternet.bnf.fr/20030329202246/http://www.univ
Accueil | Recherche initiale
Capture 11 sur 460

BnF Page archivée le : 14 avril 1997 à 02:01 GMT
Permalien : http://archivesinternet.bnf.fr/19970414020141/http://univ-lille3
Accueil | Recherche initiale
Capture 1 sur 460

Lille 3 UNIVERSITE CHARLES-DE-GAULLE

English | Index | Nouveautés | Annuaire | Laboratoires | Bibliothèque

Vous êtes à l'Université Charles-de-Gaulle - Lille 3

Nous vous souhaitons la bienvenue et vous invitons à faire la connaissance de notre Président, Monsieur Gérard LOSFELD.

- **Présentation**
  l'Université Charles-de-Gaulle - Lille 3, son histoire, la région dans laquelle elle est implantée et comment y accéder.
- **Recherche**
  Les grands axes de recherche, les laboratoires, la valorisation et la diffusion des travaux.
- **Calendrier**
  Les événements en cours ou en préparation à Lille 3 : Expositions, Colloques, Activités culturelles.
- **Formations**
  Les UFR et les Instituts, les cursus en formation initiale, continue ou en télé-enseignement

BnF

Éviter le « trou de mémoire » numérique

- Les contenus en ligne se sont multipliés :
  - Documents uniquement numériques, migrés ou à diffusion mixte

- Des solutions à trouver pour la conservation à long terme :
  - juridiques
  - techniques
  - dans la continuité des missions des bibliothèques

« Anyone who works with web archives quickly becomes used to the fact that most people have not even heard of them – even fewer understand what they are and where you might be able to access them. In 2016, however, it seemed as though web archives began to filter into the public consciousness, to move from the technology pages of the more serious newspapers to the political and even cultural sections ».

Jane Winters, A breakthrough year for web archiving in 2016?, dpconline.org, 3 february 2017.

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Plan

- Aspects juridiques, documentaires et techniques de la collecte du web français
- Les archives du Web, un enjeu patrimonial
- Modalités et outils d'accès aux archives du web
- Entre science et artisanat

Le dépôt légal en France, une mission en perpétuelle extension

- 1537 : livres
- 1648 : estampes, cartes et plans
- 1793 : partitions musicales
- 1925 : photographies, arts graphiques de toute nature
- 1938 : phonogrammes
- 1941 : affiches
- 1975 : vidéogrammes et documents multimédias
- 1977 : œuvres cinématographiques
- 1992 : documents audiovisuels de la radio télévision, édition électronique sur support (progiciels, bases de données, systèmes experts)
- **2006 : internet**

BnF

En France : une mission inscrite au code du Patrimoine

- Loi DADVSI (2006) et son décret d'application (2011) intégrés au Code du patrimoine (art. L131-2, L132-2, L132-2-1 et R132-23-1) :

  « signes, signaux, écrits, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique »

- Etablissements dépositaires :
  - INA (sites de radio et télévision)
  - BnF (tout le reste)

BnF

Une adaptation nécessaire du cadre juridique

- L'idéal d'exhaustivité est remplacé par un objectif de représentativité
- Inversion du rapport éditeur/dépositaire
  - L'obligation de dépôt légal persiste…
  - … mais n'implique pas de démarche particulière de la part des producteurs de sites
  - Procédé de collecte
- Périmètre : l'internet français, qui concerne les sites web
  - avec des adresses URL en .fr
  - dont l'auteur est une personne physique ou morale domiciliée en France
  - dont les contenus sont produits en France

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 automatisée
- Il copie les éléments qu'il trouve et qui font partie du périmètre de la collecte

Le vrai visage du robot

HERITRIX

Engine | Job Dir | Configuration | Copy Job | Scripting Console | Browse Beans

Job 23122_1492590049811

/dlweb/data/NAS730/jobs/current/g148high/23122_1492590049811/heritrix3/jobs/23122_1492590049811/crawler-beans.cxml
1 launches, last 35d3h ago

build | launch | pause | unpause | checkpoint | terminate | teardown

Job Log more

```
2017-08-24T11:32:05.587+02:00 WARNING gaip problem; using raw entity instead (in thread "ToeThread #48: dns:tiaeur-cham
2017-08-24T10:51:09.443+02:00 WARNING gaip problem; using raw entity instead (in thread "ToeThread #90: dns:startup.wp.
2017-08-24T10:39:46.422+02:00 WARNING gaip problem; using raw entity instead (in thread "ToeThread #131: http://www.del
2017-08-24T10:14:04.117+02:00 WARNING gaip problem; using raw entity instead (in thread "ToeThread #192: dns:www.shopni
2017-08-24T10:13:24.723+02:00 WARNING gaip problem; using raw entity instead (in thread "ToeThread #26: dns:www.aogua.c
```

Job is Active: RUNNING

URLs	21,944,716 downloaded + 13,614,201 queued = 35,558,932 total
Data	2.6 TiB crawled (2.6 TiB novel, 0 B dupByHash, 0 B notModified)
Alerts	3,095 tail alert log...
Rates	6.21 URIs/sec (7.26 avg); 735 KB/sec (930 avg)
Load	15 active of 200 threads; 1 congestion ratio; 5,605,219 deepest queue; 73,590 average depth
Elapsed	35d8m15s363ms
Threads	200 threads: 185 ABOUT_TO_GET_URI, 15 ABOUT_TO_BEGIN_PROCESSOR; 185 noActiveProcessor; 10 candidates, 5 fetchHttp
Frontier	RUN - 78,538 URI queues: 185 active (15 in-process; 0 ready; 170 snoozed); 0 inactive; 0 ineligible; 79 retired; 78,274 exhausted
Memory	2052037 KiB used; 3003392 KiB current heap; 3003392 KiB max heap

Reports

- CrawlSummary
- Seeds
- Hosts
- SourceTags
- Mimetypes
- ResponseCode
- Processors
- FrontierSummary
- ToeThreads

Crawl Log more

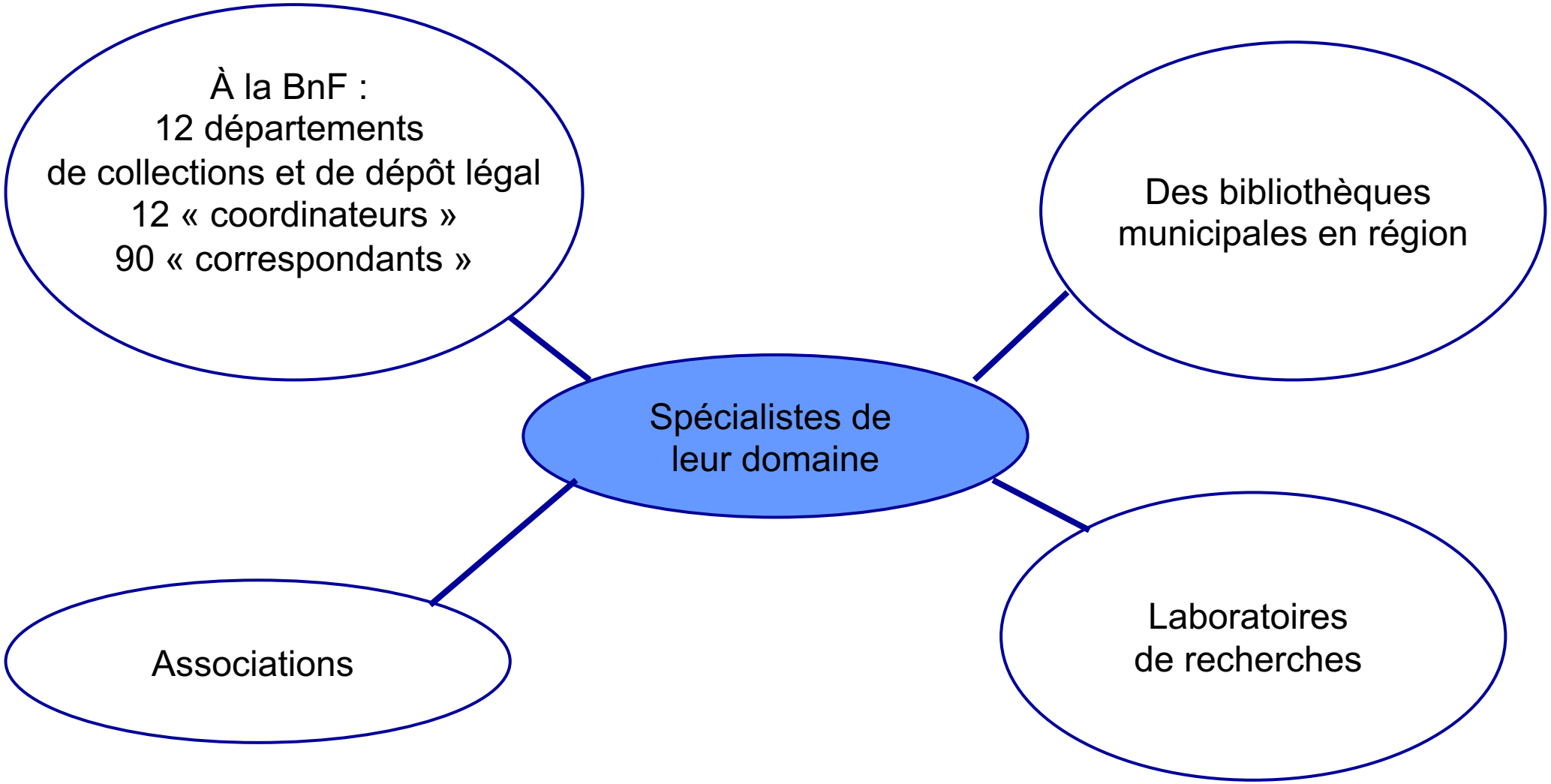
```
2017-08-24T12:04:26.087Z   200       2903 http://thumb.ccad.cnrs.fr/4904886/thumb RLLLE http://tel.archives-ouvertes.
2017-08-24T12:04:25.574Z   200       4512 http://musicimage.abcalive.com/catalog/video.contributor.62f66400-0200-11db-
2017-08-24T12:04:25.210Z -5000          - http://fr.fotolia.com/search?similar=125477738 RRLXL http://fr.fotolia.com/
2017-08-24T12:04:25.129Z   404       1246 http://www.johnsonconcrete.com/forusin.be.model LLLLXXX http://www.johnsonconc
2017-08-24T12:04:24.763Z   200       5515 http://www.avt.fr/theme/img/logo-04.png LLLLXX http://www.avt.fr/ image/png
2017-08-24T12:04:24.645Z   200       1732 http://www.truckstwork.fr/usersfiles/2502/facebook.png LLLXX http://www.truck
2017-08-24T12:04:24.642Z   404      15033 http://www.comptoir-industriel-automobile.com/test/html LLLXXX http://www.com
2017-08-24T12:04:24.330Z   200       1521 http://www.comptoir-industriel-automobile.fr/sites/all/modules/nice_menus/js/
2017-08-24T12:04:24.065Z   301          0 http://www.unisers.com/photo-voyage-img_6497.jpg-887802.html LLL http://www.
```

BnF

Constitution des collections

- Collectes larges
  - Collectes annuelles d'un très grand nombre de sites (4,5 mio en 2017)
  - Budget limité
- Collectes ciblées
  - Collectes de sites plus fréquentes et/ou plus en profondeur
  - Sélections faisant l'objet d'une politique documentaire concertée (thème, évènement)
- Des acquisitions rétrospectives d'avril/mai 1996 à 2005
- Des collections dont l'accroissement est quotidien
  - 30,9 milliards d'URL, 938,5 To de données au 31/12/2017

Qui sélectionne les sites ?



BnF

Les archives du Web, un enjeu patrimonial

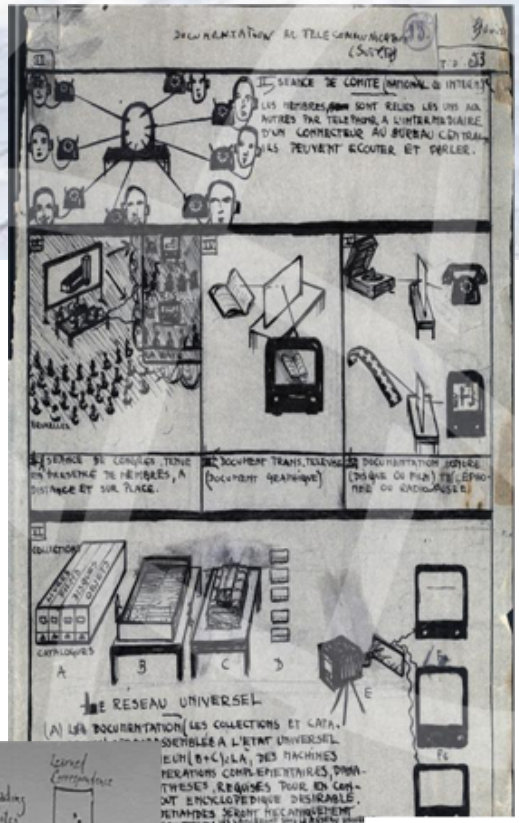
- « Le patrimoine a explosé dans tous les sens. En trente ans à peine, il a quitté le bas de laine et le « monument historique », l'église et le château pour se réfugier au lavoir du village et dans un refrain populaire. [...] Révolution silencieuse et pourtant décisive. On est passé d'un patrimoine étatique et national à un patrimoine de type social et communautaire où se déchiffre une identité de groupe; et, donc, d'un patrimoine hérité à un patrimoine revendiqué. De matériel et visible, le patrimoine est devenu invisible et symbolique, traces encore saisissables d'un passé définitivement mort, vestiges chargés d'un sens lourd, mystérieux. Bref, [...] le patrimoine a quitté son âge historique pour entrer dans son âge mémoriel : le nôtre ». Nora Pierre, «Préface», *in* Sire M-A., *La France du patrimoine, Les choix de la mémoire*, Paris, Gallimard/MONUM, 1996.

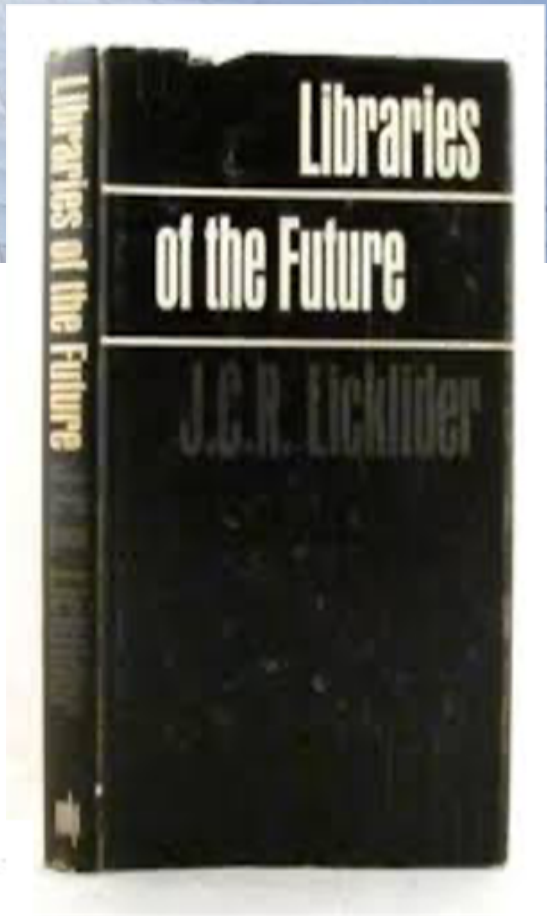
BnF

UNIVERSITÉ DU LUXEMBOURG

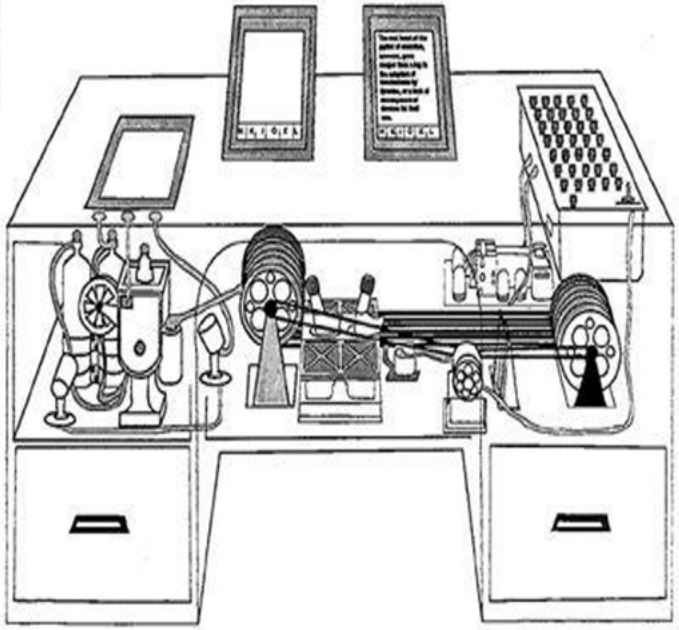
C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Paul Otlet,
Traité de
documentation
Mundaneum

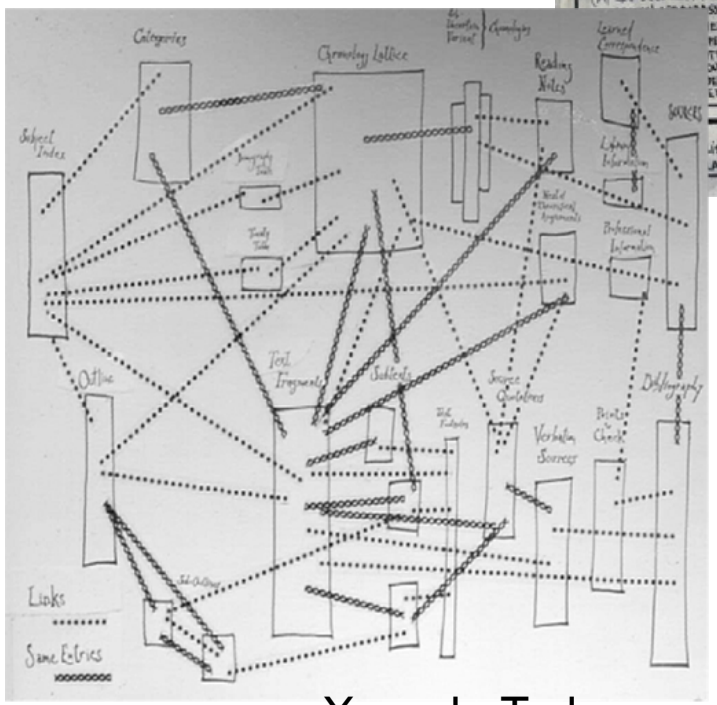




Vannevar Bush,
The Memex



Joseph Licklider,
Libraries of the
future



Xanadu Ted
Nelson

BnF

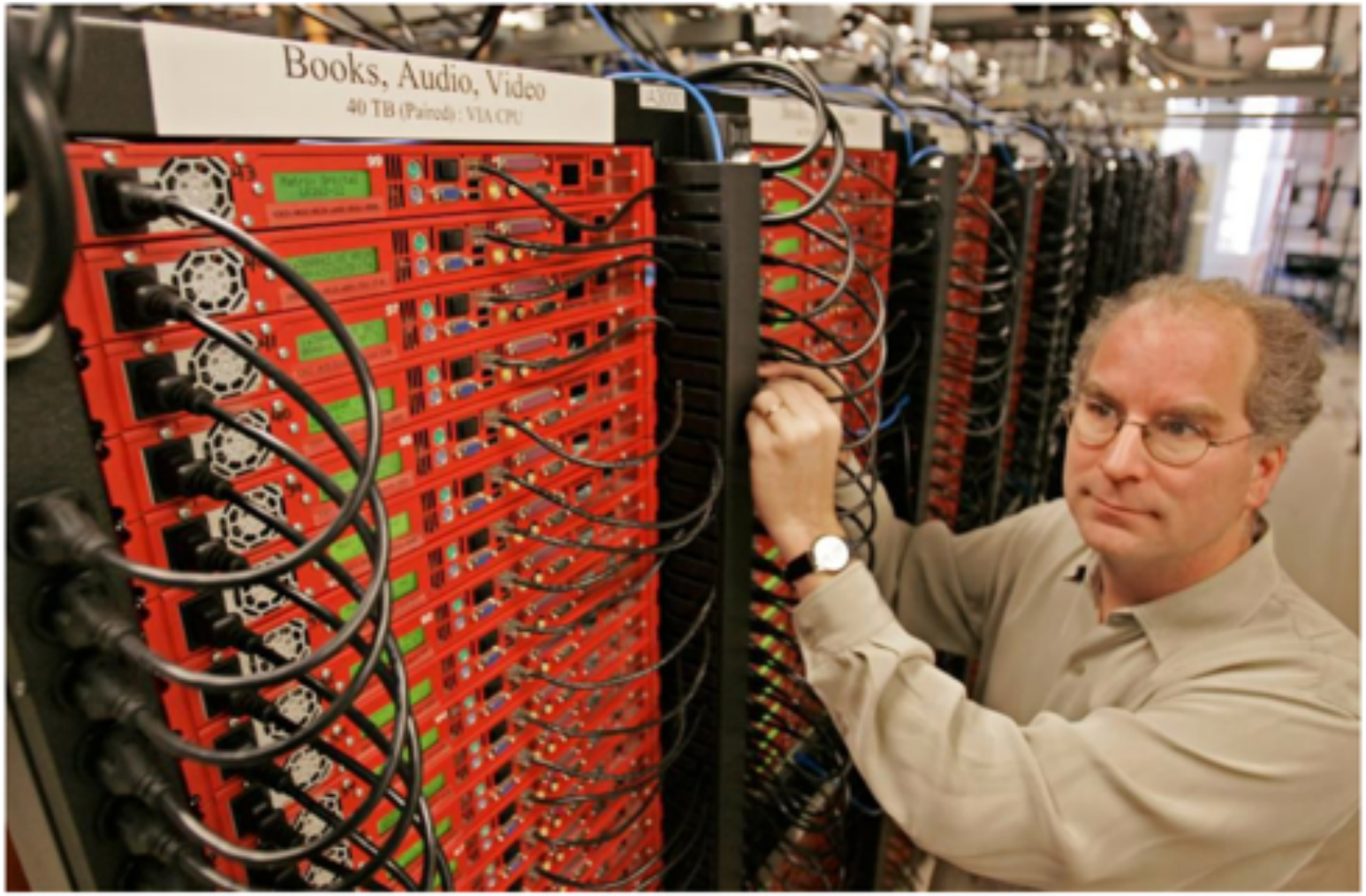
UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

JAN 18, 2016 @ 10:59 AM 4,065

2 Free Issues of Forbes

The Internet Archive Turns 20: A Behind The Scenes Look At Archiving The Web



Internet Archive founder Brewster Kahle and some of the Archive's servers in 2006. (AP Photo/Ben [+]

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Charter on the Preservation of Digital Heritage

15 October 2003

UNESDOC - (PDF) English - French - Spanish - Russian - Chinese - Arabic

The General Conference,

Considering that the disappearance of heritage in whatever form constitutes an impoverishment of the heritage of all nations,

Recalling that the Constitution of UNESCO provides that the Organization will maintain, increase and diffuse knowledge, by assuring the conservation and protection of the world's inheritance of books, works of art and monuments of history and science, that its "Information for All" Programme provides a platform for discussions and action on information policies and the safeguarding of recorded knowledge, and that its "Memory of the World" Programme aims to ensure the preservation and universal accessibility of the world's documentary heritage,

Recognizing that such resources of information and creative expression are increasingly produced, distributed, accessed and maintained in digital form, creating a new legacy – the digital heritage,

Aware that access to this heritage will offer broadened opportunities for creation, communication and sharing of knowledge among all peoples,

Understanding that this digital heritage is at risk of being lost and that its preservation for the benefit of present and future generations is an urgent issue of worldwide concern,

Proclaims the following principles and **adopts** the present Charter.

The digital heritage as a common heritage

Article 1 – Scope

The digital heritage consists of unique resources of human knowledge and expression. It embraces cultural, educational, scientific and administrative resources, as well as technical, legal, medical and other kinds of information created digitally, or converted into digital form from existing analogue resources. Where resources are "born digital", there is no other format but the digital object.

Digital materials include texts, databases, still and moving images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained.

Many of these resources have lasting value and significance, and therefore constitute a heritage that should be protected and preserved for current and future generations. This ever-growing heritage may exist in any language, in any part of the world, and in any area of human knowledge or expression.

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

HISTORY IS OUR FUTURE

And we've been trashing our history

Archive Team is a loose collective of rogue archivists, programmers, writers and loudmouths dedicated to saving our digital heritage. Since 2009 this variant force of nature has caught wind of shutdowns, shutoffs, mergers, and plain old deletions - and done our best to save the history before it's lost forever. Along the way, we've gotten attention, resistance, press and discussion, but most importantly, we've gotten the message out: **IT DOESN'T HAVE TO BE THIS WAY**.

This website is intended to be an offloading point and information depot for a number of archiving projects, all related to saving websites or data that is in danger of being lost. Besides serving as a hub for team-based pulling down and mirroring of data, this site will provide advice on managing your own data and rescuing it from the brink of destruction.

Currently Active Projects (Get Involved Here!)

Archive Team recruiting

- Want to code for Archive Team? Here's a starting point.

Warrior based projects

- Google Baraza: Read-only, getting deleted on April 30, 2015. **IRC Channel #bonanza**.
- Last.fm: Switching codebases in the first two weeks of April, some user-generated content might be lost. **IRC Channel #lastchance.fm**.
- Trovebox: Closing March 31, 2015. **IRC Channel #treasuretrove**.
- Blogger: Google was to restrict public access to adult blogs, but they backed down. We're backing up Blogger anyway. **IRC Channel #frogger**.
- URLTeam: URL shorteners were a fucking awful idea. **IRC Channel #urlteam**.
- TwitPic: allegedly acquired and possibly to be rearranged by Twitter. We're saving it in its current state. **IRC Channel #quitpic**.
- Halo: Downloading all the read-only data from Halo. **IRC Channel #yolohalo**.

Help us: **Download and run your warrior**.
What's on: online tracker.
Combined project activity graphs here.

http://archiveteam.org/index.php?title=Main_Page

FAQ sur l'archivage web

Généralités
Collection
Collecter et archiver
Utilisation

Généralités

POURQUOI LA BIBLIOTHEQUE NATIONALE SUISSE ARCHIVE-T-ELLE LE WEB ?
La Bibliothèque nationale suisse a pour mandat de collectionner, d'archiver et de rendre accessible le patrimoine culturel. Ce mandat inclut également la sauvegarde du savoir numérique pour les générations futures. Dans le cadre des Archives Web Suisse qu'elle constitue en collaboration avec les bibliothèques cantonales suisses et d'autres bibliothèques spécialisées, la Bibliothèque nationale suisse se fixe pour objectif de conserver des parties de l'"internet suisse".

QUELLES STRATEGIES LA BIBLIOTHEQUE NATIONALE SUISSE APPLIQUE-T-ELLE POUR L'ARCHIVAGE DU WEB ?
La Bibliothèque nationale suisse a opté pour une stratégie sélective. Elle met l'accent sur les sites web patrimoniaux qui ont un fort lien avec la Suisse et qui sont accessibles librement : sites web sur les cantons et les communes, domaines spécifiques tels que sciences sociales ou littérature suisse. Ce sont avant tout les bibliothèques cantonales et d'autres bibliothèques spécialisées qui sélectionnent les sites web. Des collections spéciales sont constituées à l'occasion de manifestations spéciales qui ont lieu en Suisse : les sites web concernant un évènement comme les élections fédérales de 2011 sont alors collectés et archivés. La sélection de sites web permet de créer des instantanés du web suisse et de les enregistrer pour la postérité. Les critères de sélection sont décrits en détail dans la notice Collecte. La Bibliothèque nationale renonce pour l'instant à un harvesting du domaine .ch.
Notice Collecte

ARQUIVO.PT

Type word or URL | Search

Advanced search

Search pages from the past

Millions of contents archived since 1996

Meet the service

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Do Web Archives have Politics?

The MIT Press

Do Artifacts Have Politics?
Author(s): Langdon Winner
Source: *Daedalus*, Vol. 109, No. 1, Modern Technology: Problem or Opportunity? (Winter, 1980), pp. 121-136
Published by: The MIT Press on behalf of American Academy of Arts & Sciences
Stable URL: http://www.jstor.org/stable/20024652
Accessed: 06/10/2009 20:50

Bulletin of the
Seismological Society of America

Vol. 64 October 1974 No. 5

IS THE SEQUENCE OF EARTHQUAKES IN SOUTHERN CALIFORNIA, WITH AFTERSHOCKS REMOVED, POISSONIAN?

By J. K. Gardner and L. Knopoff

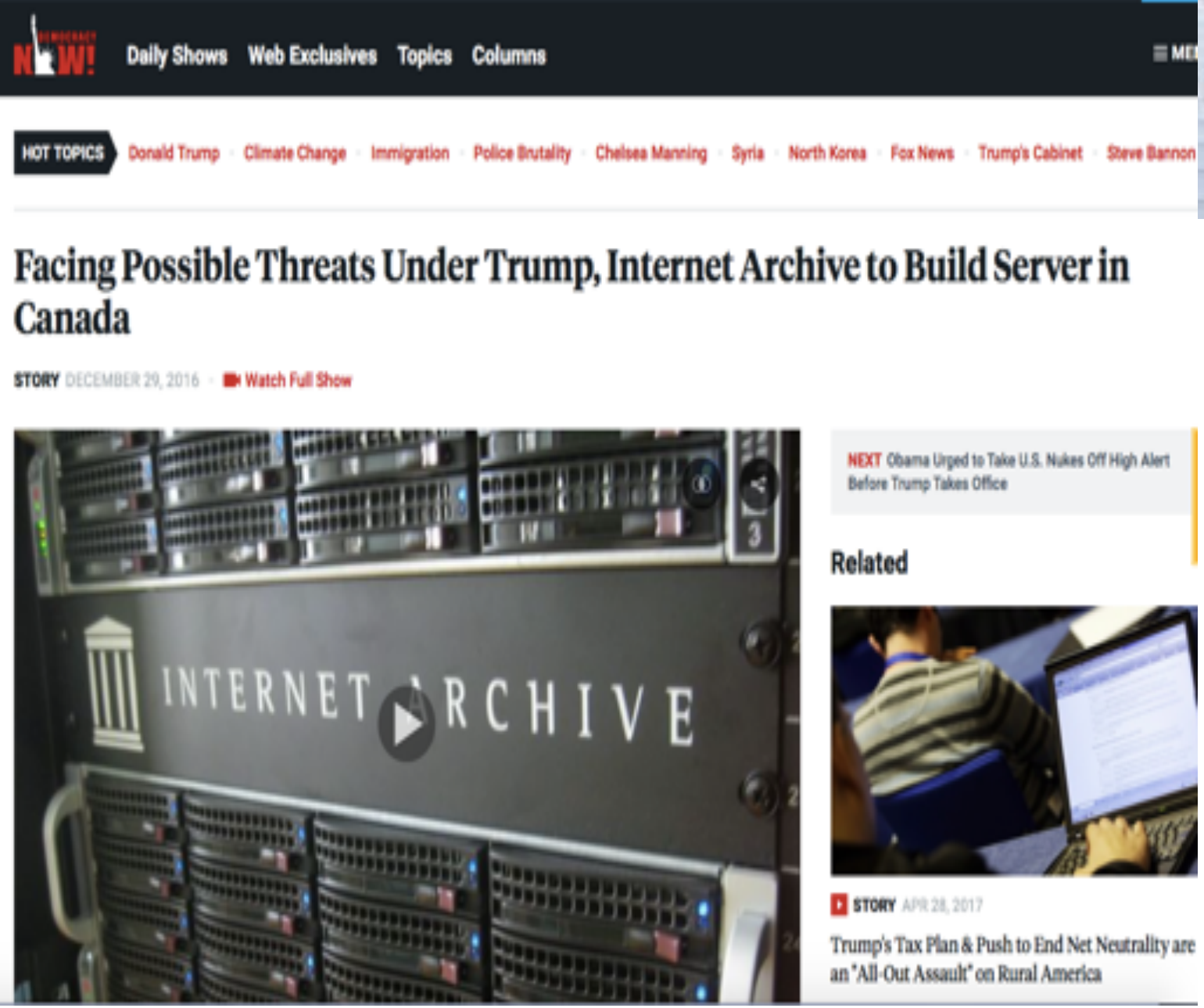
Abstract

Yes.

{BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY



https://www.democracynow.org/2016/12/29/facing_possible_threats_under_trump_internet

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

theguardian

UK world sport football opinion culture business lifestyle fashion environment tech travel all sections

home › opinion columnists letters editorials

Internet
You told us

The Conservatives' website purge has destroyed history

Alex Hern

The Tory party went beyond the acceptable into an Orwellian realm when it wiped a decade of speeches from the internet

243 311

Contact author

@alexhern

Thursday 14 November 2013 13.39 GMT

'Remember the 2007 speech in which George Osborne committed to '2% a year spending rises for the next three years'? Well, the internet doesn't any more.' Photograph: Chris Ison/PA

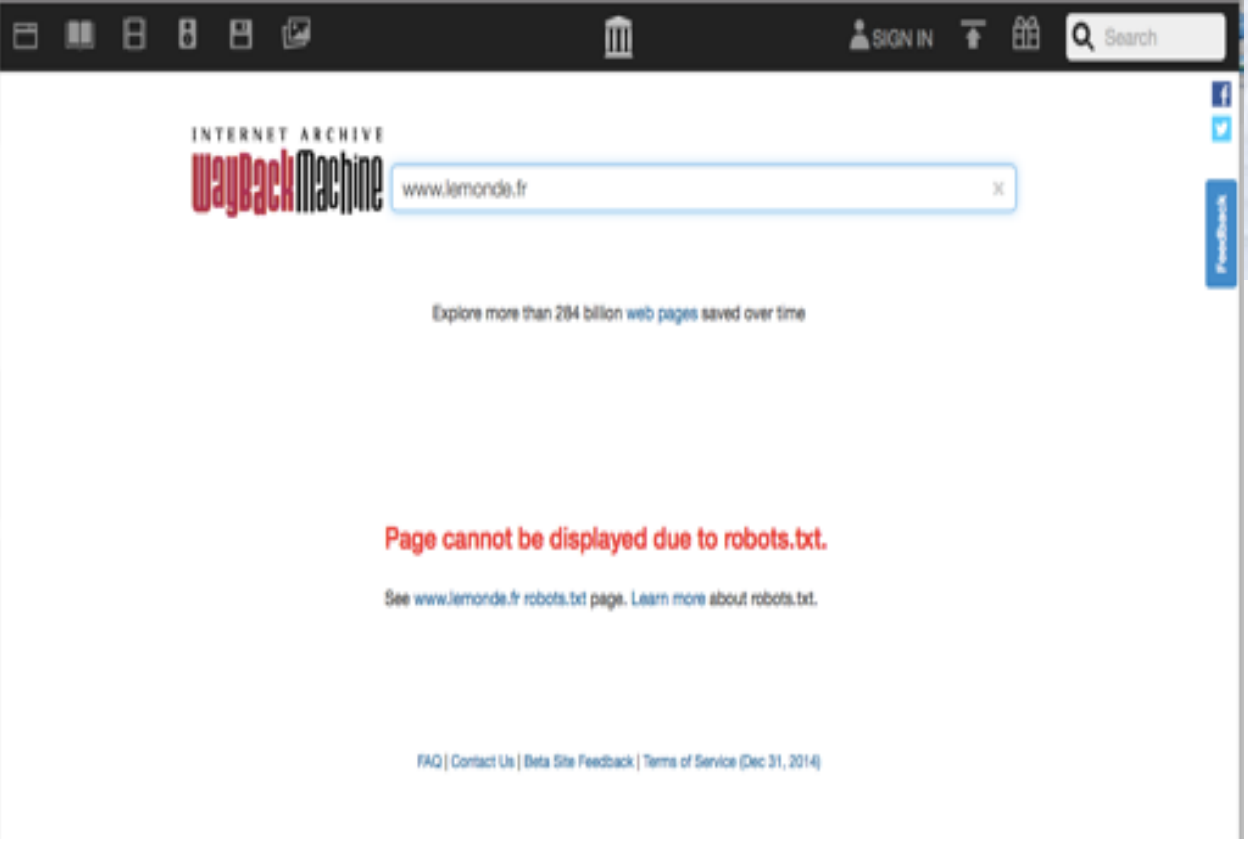
The Tory party has chucked its history in the dustbin. In the course of a website redesign, the Conservatives deleted their digital archive of political speeches from 2000 to 2010, so that everything they said until the coalition government formed is gone.

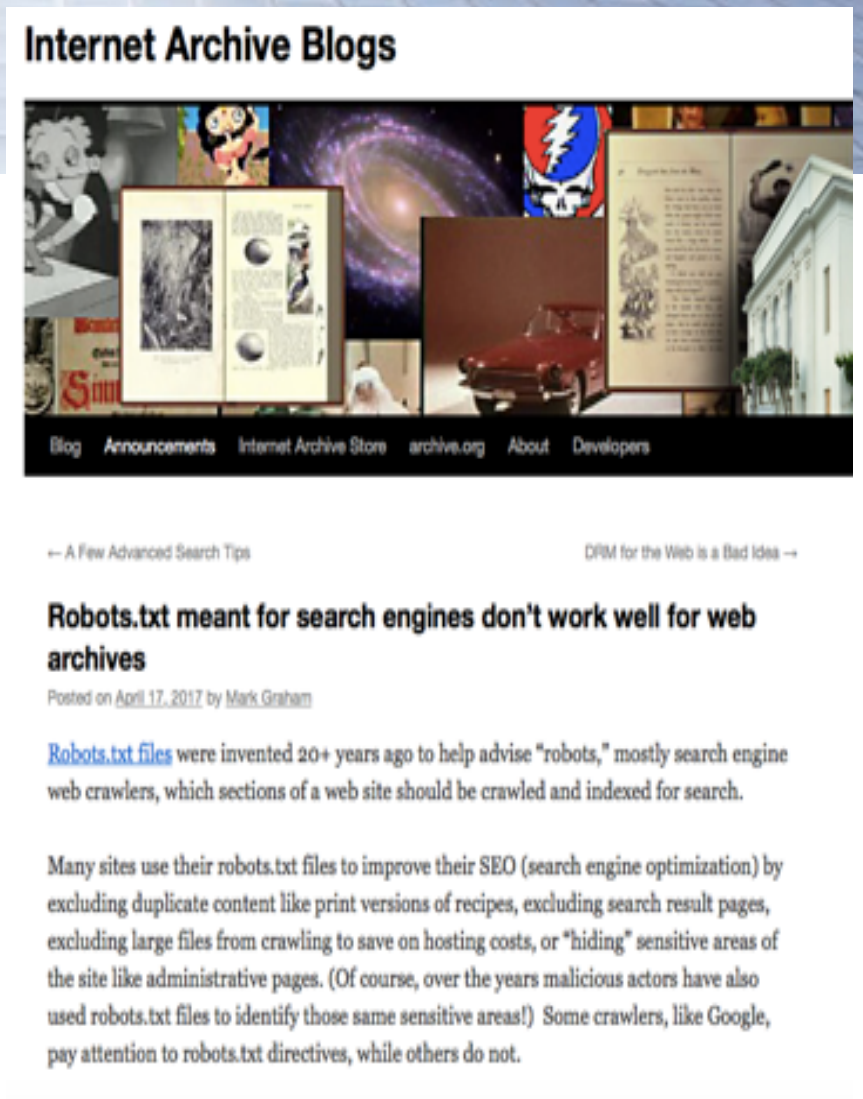
https://www.theguardian.com/commentisfree/2013/nov/14/conservatives-website-purge-history-tory-party-speeches-internet

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY





https://blog.archive.org/2017/04/17/robots-txt-meant-for-search-engines-dont-work-
well-for-web-archives/

BnF

UNIVERSITÉ DU LUXEMBOURG

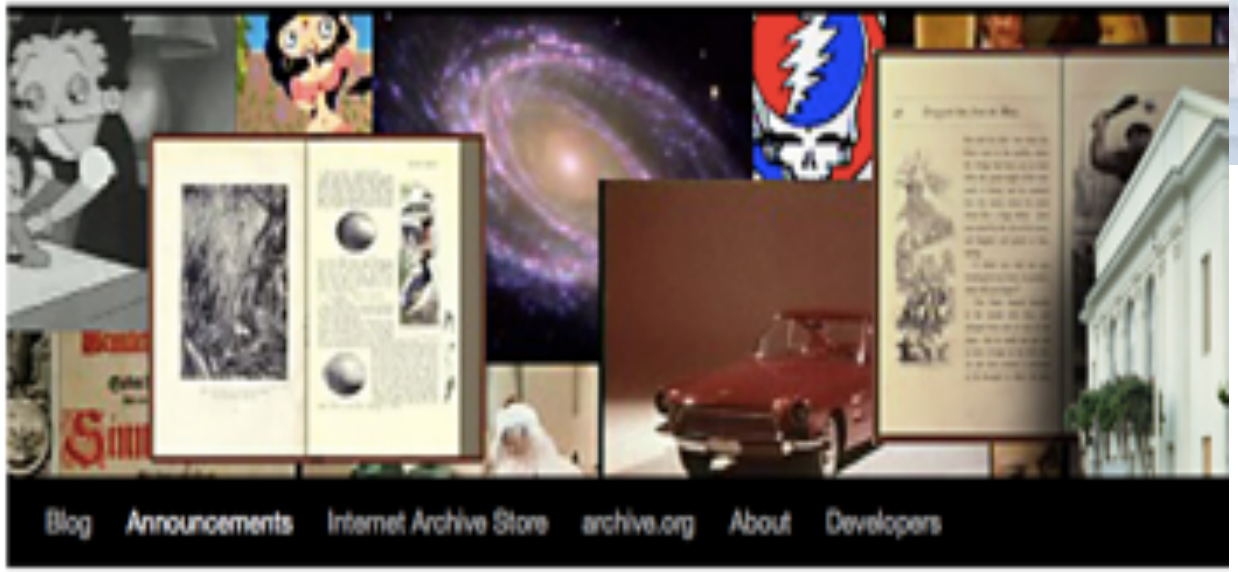
C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

DocNow

DocNow is a tool and a community developed around supporting the ethical collection, use, and preservation of social media content.

Photo by Neil Cooler

Internet Archive Blogs



https://blog.archive.org/2017/04/11/who-blocked-the-archive-in-jordan/

← TV News highlights: stimulus package, global food demand, carbon emissions, and more

TV News highlights: Hitler, Syria, NATO, and more →

Who Blocked the Archive in Jordan?

Posted on April 11, 2017 by Chris Butler

A new article at the Jordanian news site 7iber.com by Reem Al-Masri tells the story of how the Archive was blocked in the country of Jordan last year and came to be unblocked. Al-Masri highlights many of the problems inherent in the general Jordanian legal approach to censoring sites. Indeed, government censorship is an extremely worrisome issue in and of itself. But, the specific instance of archive.org's block is all the more extraordinary because it appears to have been enacted outside of the established Jordanian legal process.

Jordanian law currently empowers a single body to issue a block order, the Jordanian Media Commission. Based on evidence from the University of Toronto's Citizen Lab and various user reports, it's clear that archive.org was blocked some time in 2016. However

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Multi-stakeholderism

INTERNATIONAL INTERNET PRESERVATION CONSORTIUM

Groupes

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Les modalités de consultation déterminées par la loi

- La consultation s'effectue :
  - À la **BnF** et dans **tout organisme habilité** à mettre en œuvre cette consultation par arrêté
  - Sur des postes individuels équipés d'interfaces d'accès, de recherche et de traitement fournies par la BnF et dont l'usage est strictement réservé à des chercheurs dûment accrédités
- À la BnF : accès selon les mêmes principes appliqués aux autres collections de dépôt légal (espaces Recherche)

Pour un accès distant aux archives du web

Où consulter les archives de l'internet de la BnF ?

Filtrer...

Accès à distance

- Archives de la Martinique - Fort-de-France
- Archives Départementales de la Guadeloupe - Gourbeyre
- Bibliothèque d'Étude et conservation - Besançon
- Bibliothèque d'Étude et du Patrimoine - Toulouse
- Bibliothèque du Patrimoine - Clermont-Ferrand
- Bibliothèque Francophone et Multimédias - Limoges
- Bibliothèque Mériadeck - Bordeaux
- Bibliothèque municipale Part-Dieu - Lyon
- Bibliothèque Nationale Universitaire - Strasbourg
- Bibliothèque patrimoniale et d'étude - Dijon
- Bibliothèque patrimoniale Villon - Rouen
- Bibliothèque Stanislas - Nancy
- Bibliothèque Toussaint - Angers
- Médiathèque centrale Emile Zola - Montpellier
- Médiathèque François-Mitterand - Poitiers
- Médiathèque Jean-Levy - Lille

Accès BnF

- BnF - François-Mitterand
- BnF - Maison Jean Vilar - Avignon
- BnF - Site Richelieu-Louvois

BnF

Un site, des captures

- Un grand principe : des captures de sites internet proposées selon des entrées chronologiques

- Une interface, plusieurs modes de consultation :
  - Recherche par URL, sur toutes les collections
  - Parcours guidés

- Un logiciel : OpenWayback

BnF

BnF Archives de l'internet

A propos | Aide

Remonter le temps

Trouver l'archive d'un site, une page, un fichier en indiquant son adresse URL

http://

Rechercher l'URL

» Recherche avancée

Recherche par mot dans le web français des années 1996-2000

Tester le prototype

Découvrir les collections à travers des sélections thématiques

PRESSE ET ACTUALITE

LE WEB GOURMAND | LE WEB DES ANNÉES 1... | LA RONDE DES FESTIV... | LE MARCHÉ DE L'ART | JEAN-PHILIPPE RAMI...

Tous les parcours guidés

Suivez l'expédition Tara Oceans dans l'infiniment petit du peuple de la mer.

Archive du 17 juillet 2012 à 12:19

Département Sciences et techniques

La science en BD, pour mourir moins bête.

Archive du 01 avril 2012 à 10:30

Département Sciences et techniques

Revivez le colloque de la BnF sur les journaux intimes à travers les dessins de Michel Longuet.

Archive du 27 août 2007 à 10:21

Département Littérature et art

Cliquez et entrez dans Ge-nove, le labyrinthe de mots et d'images élaboré par Benoît Vincent autour de la ville de Gênes.

Archive du 19 mars 2015 à 21:23

Département Littérature et art

Éric Chevillard

BnF

BnF Archives de l'internet

A propos | Aide

http://lemonde.fr

Rechercher l'URL

» Recherche avancée

Parcours guidés

» Presse et actualité

Recherche par URL

- URL : http://lemonde.fr
- Date : du 1 jan. 1996 au 11 jan. 2017

9 209 résultats

» Modifier la recherche

Tout ouvrir | Tout fermer

2017	20 résultats
2016	1 232 résultats
2015	1 117 résultats
2014	1 423 résultats
2013	1 784 résultats
2012	1 679 résultats

BnF

BnF
Page archivée le : 06 mai 2012 à 08:05 GMT
Permalien : http://archivewebtemp.bnf.fr/20120506080521/http://www.lemo
Accueil | Recherche intégrée
Capture 2 247 sur 9 228

Le Monde.fr

Le second tour de la présidentielle en direct

Les bureaux de vote en métropole ont ouvert à 8 heures. Une majorité de Français d'outre-mer et de l'étranger a elle déjà voté.

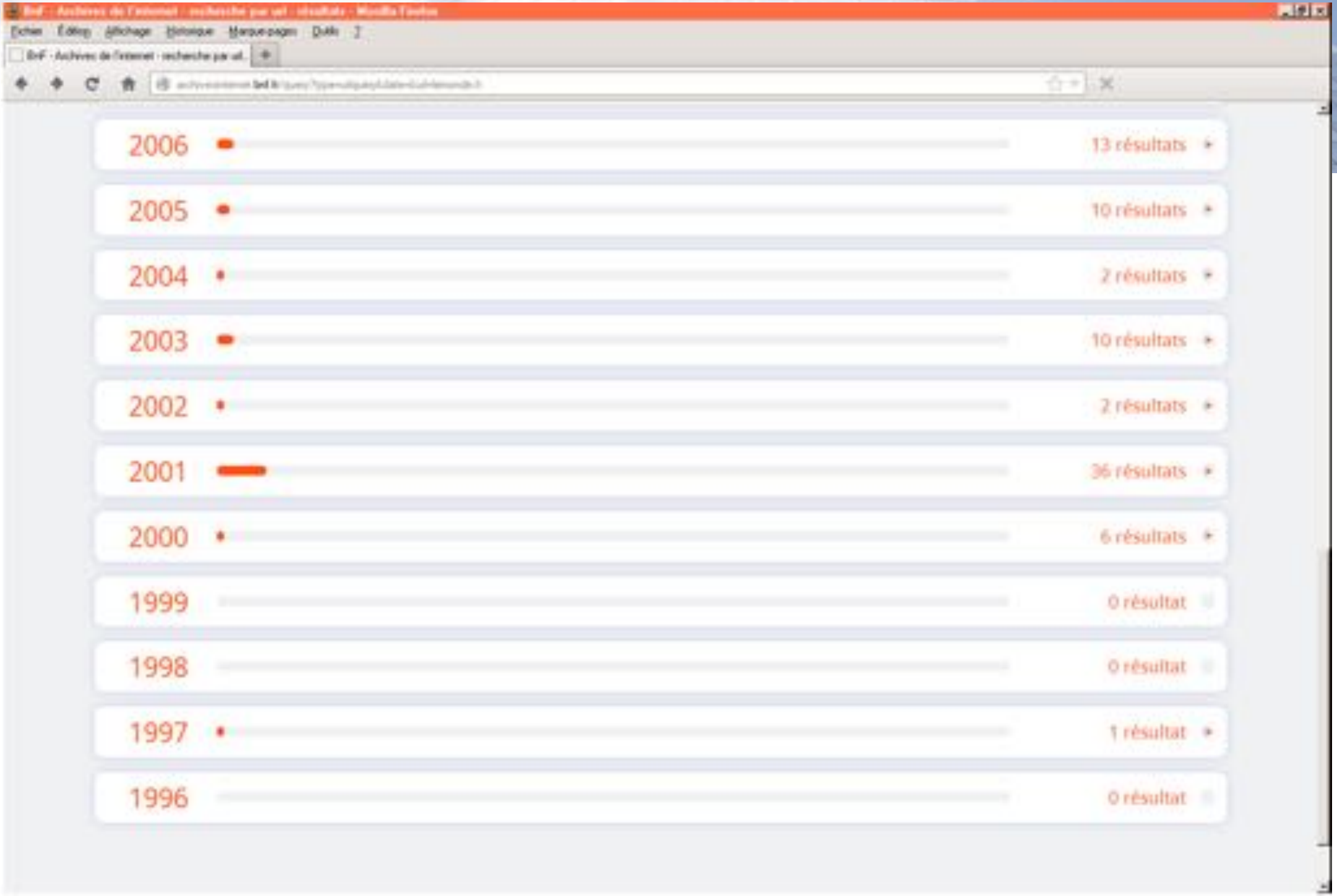
Suivez en direct la journée électorale.

08:53 La vote a commencé en métropole

BnF

(BnF
Le Monde.fr
INTERNATIONAL POLITIQUE SOCIÉTÉ ÉCONOMIE CULTURE IDÉES SPORT SCIENCES TECHNO STYLE VOUS
ÉDITION ABONNÉS
EN CE MOMENT Présidentielle 2012 François Hollande Nicolas Sarkozy Grèce Serbie
François Hollande élu président de la République
Le candidat du PS a devancé Nicolas Sarkozy au second tour de l'élection présidentielle, avec 51,9 % des voix contre 48,1 % au président sortant, selon les estimations Ipsos pour "Le Monde", France télévisions et Radio France à partir des premiers bulletins dépouillés.
FRANÇOIS HOLLANDE
51,9 %
NICOLAS SARKOZY
48,1 %
20:02 Dans le bureau de François Hollande 10 minutes avant les résultats

(BnF


2006
13 résultats
2005
10 résultats
2004
2 résultats
2003
10 résultats
2002
2 résultats
2001
36 résultats
2000
6 résultats
1999
0 résultat
1998
0 résultat
1997
1 résultat
1996
0 résultat

BnF

Bienvenue sur Le Monde

BnF

Page archivée le : 22 juin 1997 à 15:21 GMT

Permalien

Accueil | Recherche indexée

Capture 1 sur 3,228

BIENVENUE sur la toile du MONDE

Globe Online

En raison de la non-distribution du journal

L'accès au journal du 19 juin, daté 20, est gratuit

Les résultats des Grandes Ecoles

Les élections Législatives 1997

La fête de la Musique

Le Monde

Site en Europe / Site aux Etats-Unis

La *Une* du Monde | Le *multimedia* | Le *journal* complet

BnF

Le Monde - Législatives 97 - Mozilla Firefox

Fichier Éditer Affichage Historique Marque-pages Outils ?

Le Monde - Législatives 97

www.lemonde.fr/elections/index.html

BnF Archives de l'internet

A propos | Aide

http://www.lemonde.fr/elections/framebs.html Rechercher l'URL » Recherche avancée Parcours guidés » Presse et actualité

BnF Archives de l'internet

A propos | Aide

http://www.lemonde.fr/elections/framebs.html Rechercher l'URL » Recherche avancée Parcours guidés » Presse et actualité

Recherche par URL

- URL : http://www.lemonde.fr/elections/framebs.html

Cette page n'est pas dans les archives.

» Lancer une nouvelle recherche

» Revenir à la page précédente

Informations légales et droits de reproduction | Contact | 6.3.3-BnF (g462)

Copie et utilisation des archives

- Possibilité de conserver et citer des références (permalien)
- Possibilité de copier des extraits de texte à des fins de citations
- Interdiction de télécharger des contenus

Nouvelles approches, nouveaux services

BnF API et jeux de données

Découvrez et utilisez les données de la BnF

COMPRENDRE

Le portail BnF API et jeux de don
de récupérer les métadonnées
collections numérisées de la Br
(images et textes, métadonnées

Plusieurs formats, plusieurs te
de la bibliothèque : alimentation
datavisualisation, etc.

Développeurs et développeuses,
chaîne du livre, *digital humanists*,
diffusées par la BnF n'attenden

À propos de BnF API et jeux de données

Actualités

BnF Archives de l'internet Labs

Accueil | Collections | Services | Aide

Rechercher

» Recherche avancée
» Recherche experte

» Se connecter

Archives de l'internet Labs permet d'explorer une partie des collections issues du dépôt légal de l'internet, constituées depuis 1996, en utilisant la recherche plein texte. *Archives de l'internet Labs* est complémentaire de l'application *Archives de l'internet* qui permet de faire des recherches par adresse URL et par date, de visualiser les pages archivées et de naviguer dans les archives du web français.

Des services complémentaires sont également proposés aux chercheurs dont les travaux sont réalisés sous couvert d'une convention. Trois collections sont actuellement disponibles.

Actualités 2010-2017

Attentats 2015

Incunables 1996-2000

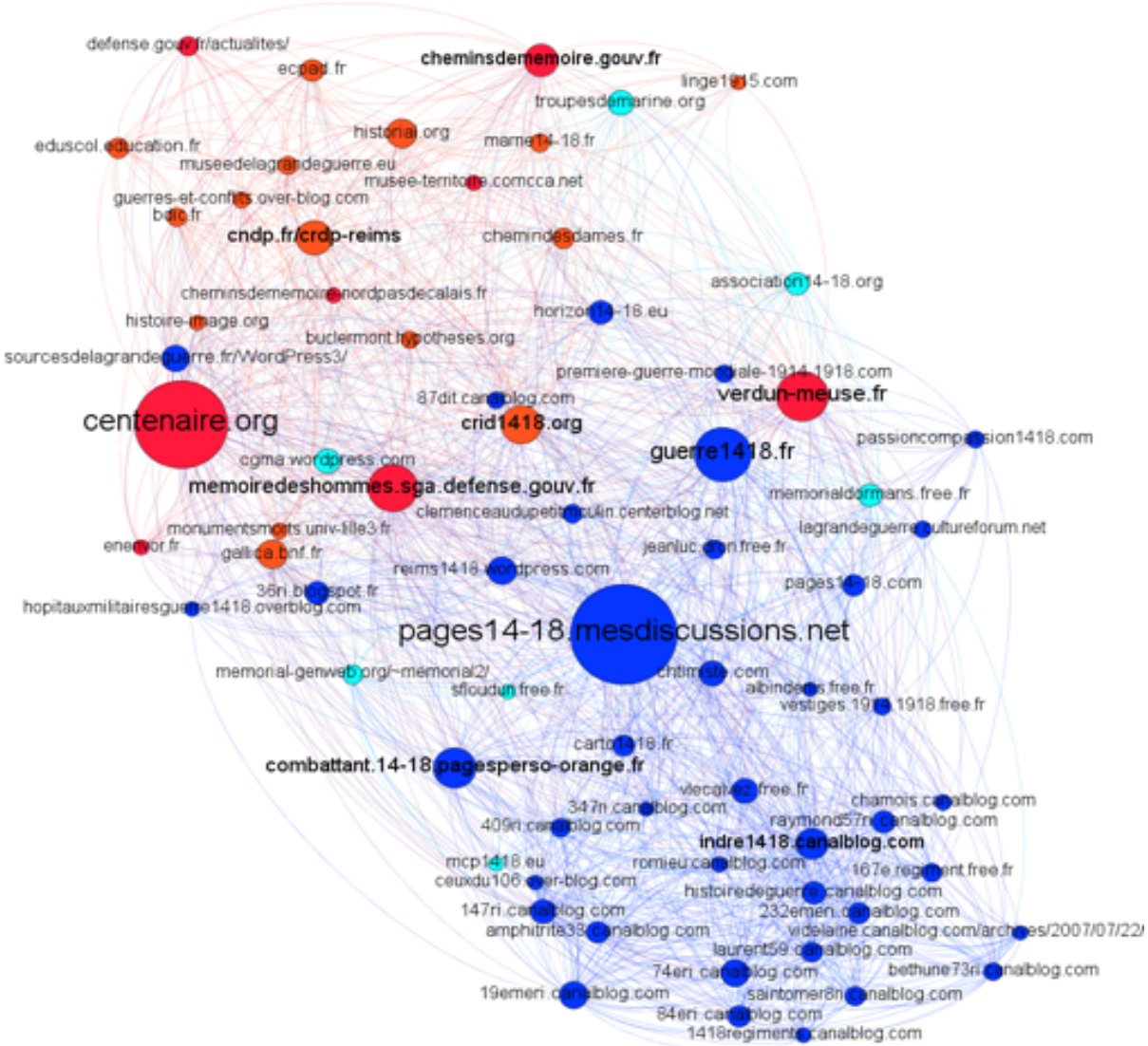
BnF

Un intérêt croissant pour les humanités numériques

- Fouille de données et de textes, analyse de liens…
- Accueil sur site, fourniture d'accès aux données et aux métadonnées, accompagnement des équipes de recherche, (co)développements d'outils et de techniques (projets conventionnés)
- Création d'un « Laboratoire d'étude et d'analyse de corpus de corpus numériques » (en cours)

Le devenir du patrimoine numérise en ligne : l'exemple de la Grande guerre

- 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* (Investissements d'avenir, réf. ANR-11-LABX-0026-01) : https://hal.archives-ouvertes.fr/hal-01425600



BnF

Web90

- Web90, projet ISCC/CNRS avec financement ANR sur « Patrimoine, mémoires et histoire du web dans les années 1990 » : http://web90.hypotheses.org/

BnF

Entre science et artisanat …

@Ray Clid
[*Un éléphant qui se balançait*
Sur une toile toile toile d'araignée.
C'était un jeu tellement tellement amusant que,
Tout à coup : Badaboum!]

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

« The most significant barrier to working with web archives is, quite simply, that it is difficult; it requires skills that many historians do not have, and in the short term may be unwilling to learn; it involves acknowledging a degree of ignorance with which otherwise seasoned researchers may be uncomfortable ».

Winters Jane (2017) ‘Breaking in to the mainstream: demonstrating the value of internet (and web) histories’, *Internet Histories*, Issue 1-2, pp. 173-179.

BnF

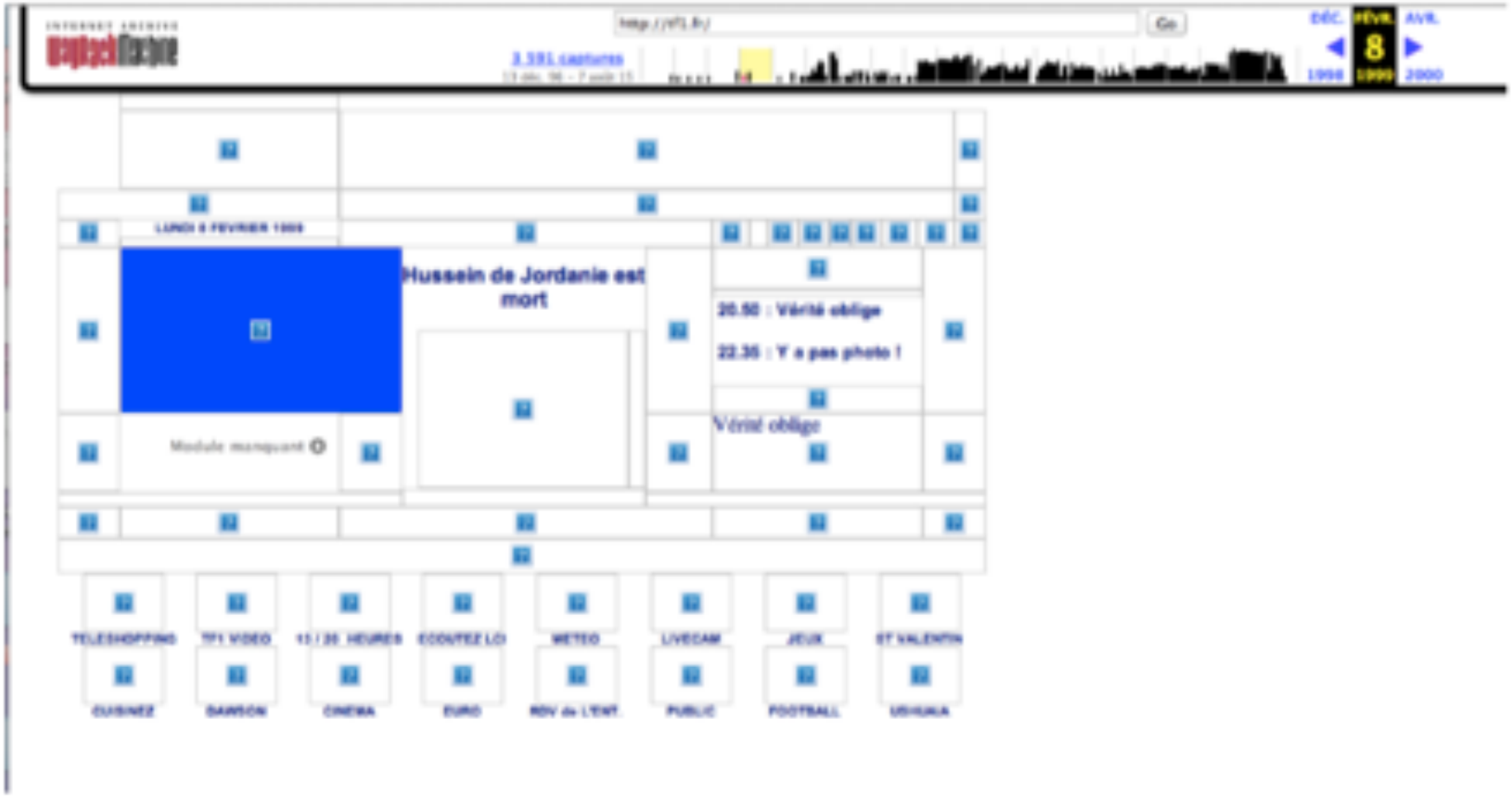
UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY



Site tf1.fr archivé au 8 février 1999 par Internet Archive
https://web.archive.org/web/19990208010527/http://tf1.fr/

BnF

UNIVERSITÉ DU LUXEMBOURG

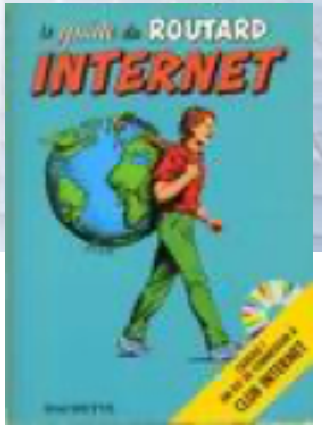
C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Page du site de l'Aftel conservée par Internet Archive le 25 janvier 1997
https://web.archive.org/web/19970125032641/http://www.aftel.fr/ssacti.htm

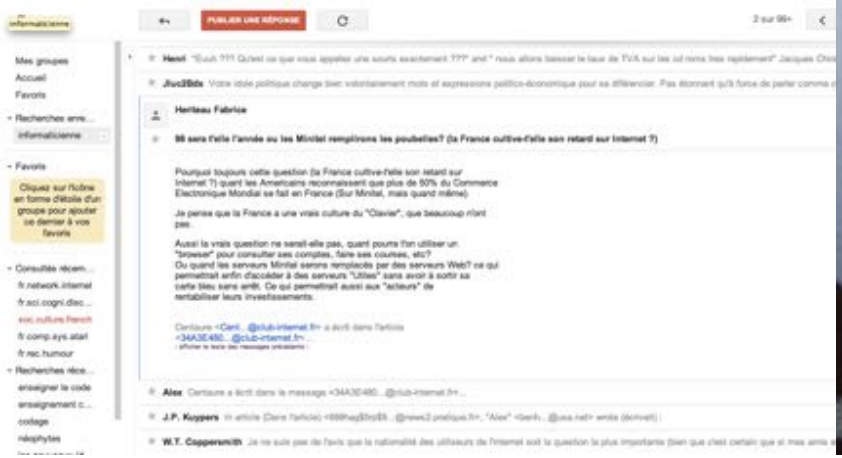
BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY



http://www.letenneur.com



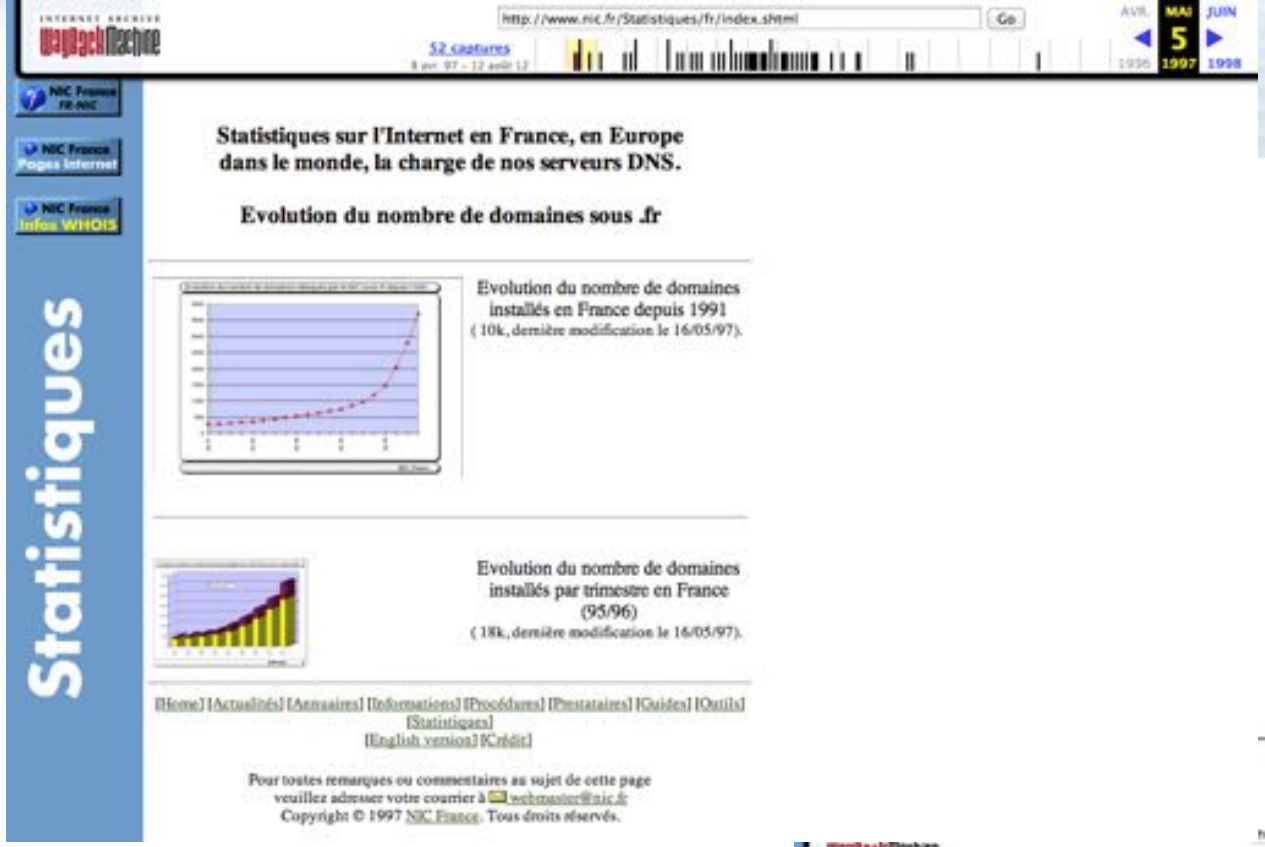
Entretien avec Stéphane Bortzmeyer – 1ère partie

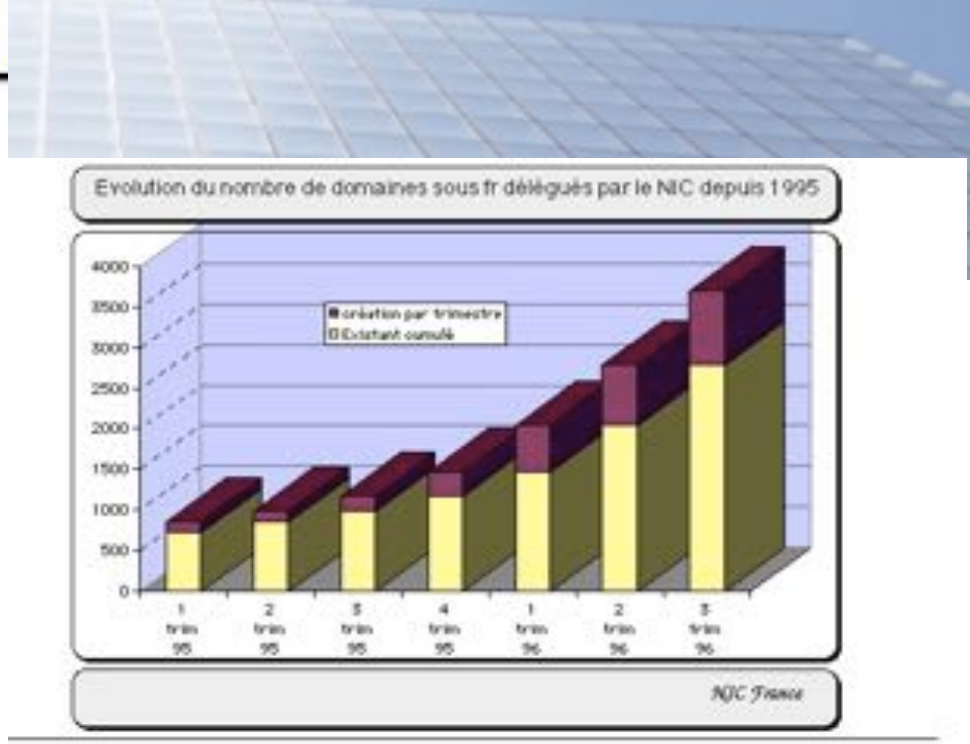
Le 23 avril 2015, Stéphane Bortzmeyer était notre premier grand témoin interviewé dans le cadre du projet Web90 (par Camille Paloque-Berges et Valérie Schafer) pour revenir sur sa découverte et son expérience du Web des années 90.

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY





NIC France
NETWORK INFORMATION CENTER
MINITEL : 3615 NICFRANCE

nic.fr

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

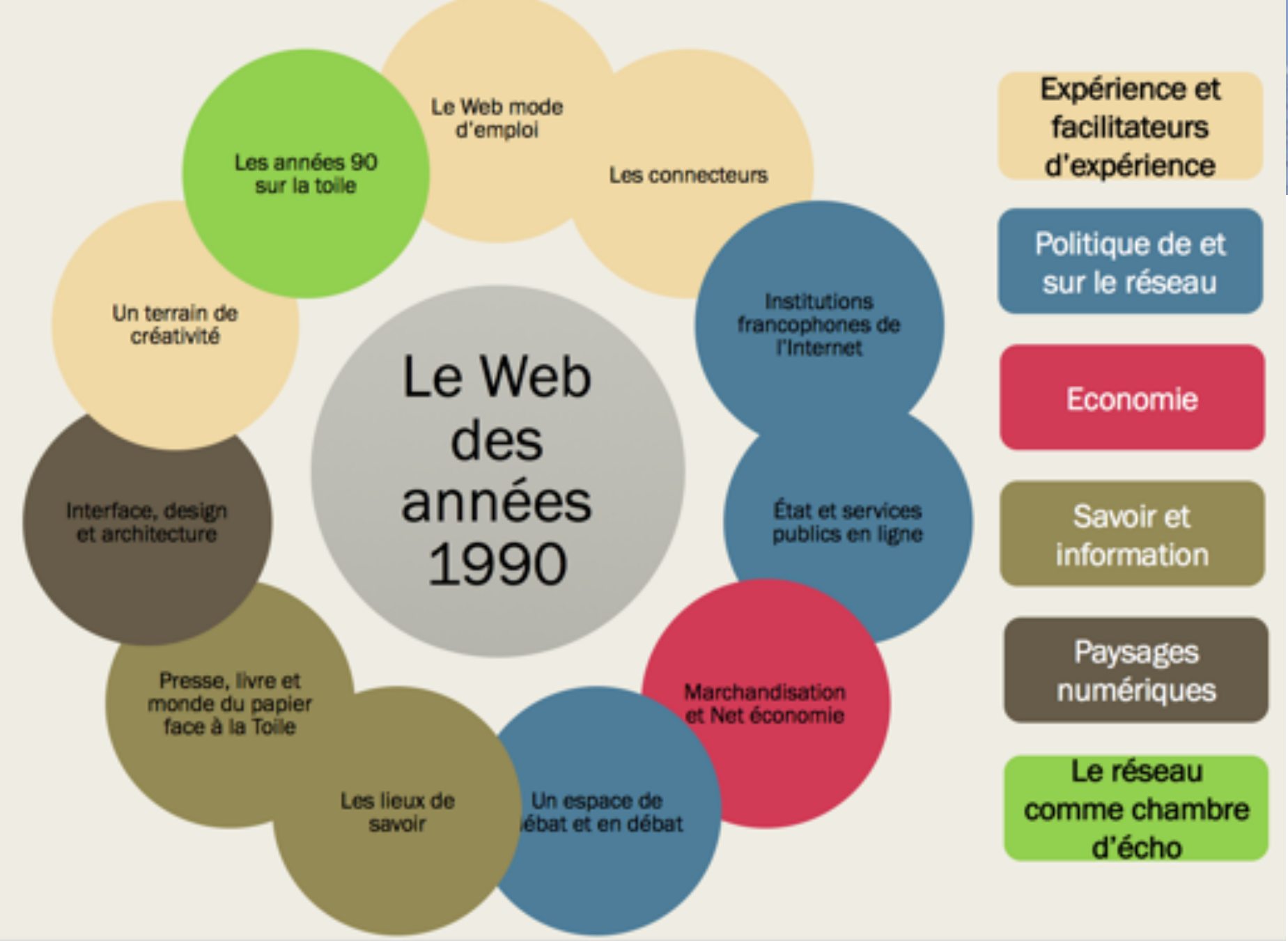
Internet Archive, Homepage of the CCI of Amiens, 12 April 1997
https://web.archive.org/web/19970412025333/http://www.amiens.cci.fr/

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH
LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Parcours guidé
BnF

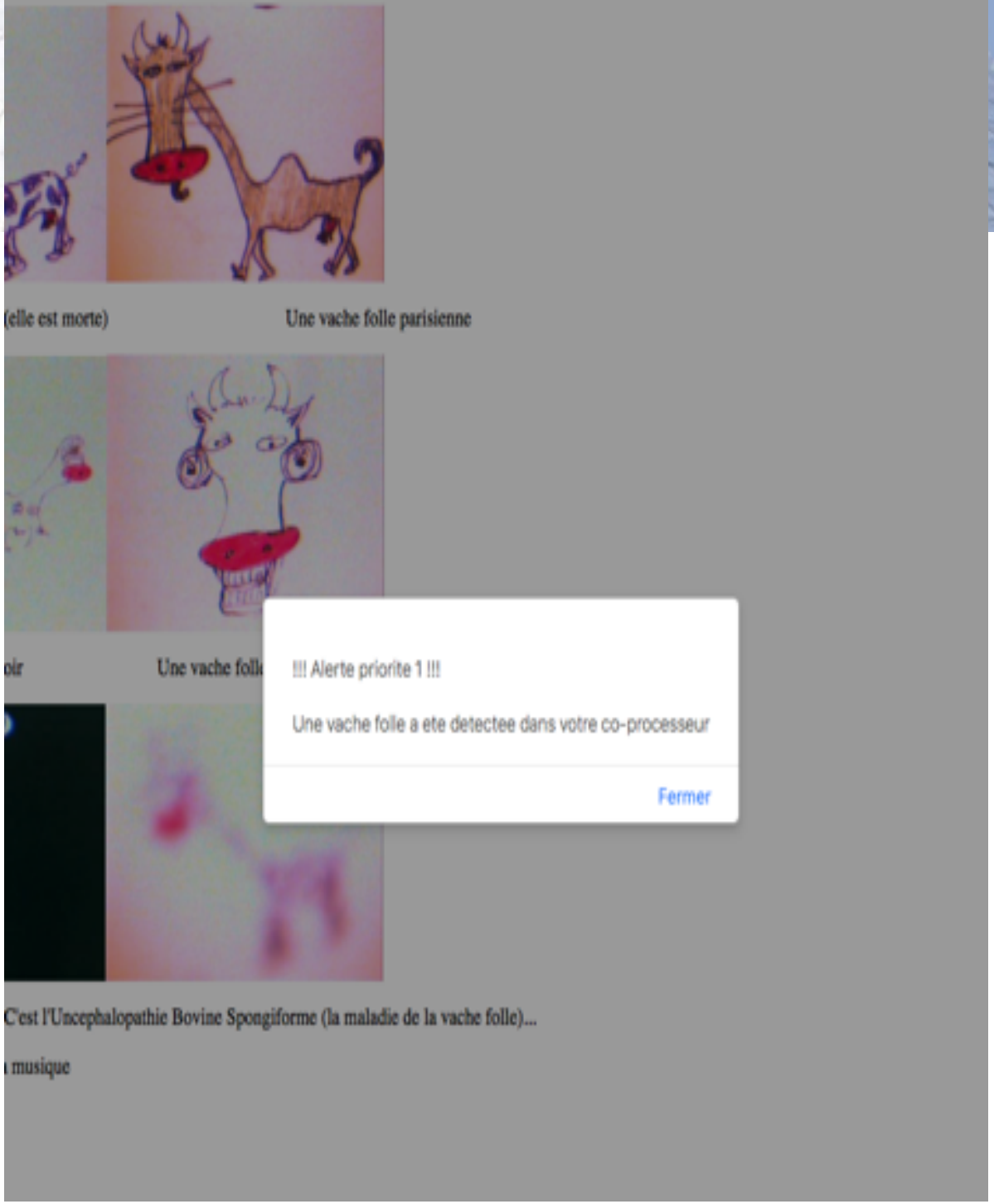


BnF

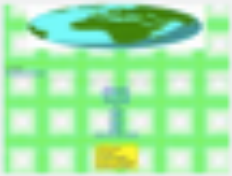
UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

BnF
Page archivée le : 06 mai 1997 à 10:52 GMT
Permalien : http://val100archivesinternet.bnf.fr/19970506105222/http://ble
Il faut bien s'occuper pendant les cours:
Une vache folle bretonne (en sabots)
Une vache folle batarde
Une vache folle anglaise (elle est morte)
Une vache folle parisienne
Une vache folle dans le noir
Une vache folle bretonne dans le brouillard anglais
Vous connaissez l'UBS ? C'est l'Uncephalopathie Bovine Spongiforme (la maladie de la vache folle)...


(elle est morte)
Une vache folle parisienne
!!! Alerte priorite 1 !!!
Une vache folle a ete detectee dans votre co-processeur
Fermer
C'est l'Uncephalopathie Bovine Spongiforme (la maladie de la vache folle)...

BnF



Le rouleau avant la page

http://www.polytechnique.fr

Les pages personnelles dans les années 1990 sont souvent marquées par deux caractéristiques : le mélange des sujets et l'aspect de « rouleau », c'est-à-dire de page unique qu'il faut faire défiler plutôt qu'un ensemble de pages au sein desquelles il faut naviguer.

Archive du 04 avril 1997 à 02:38 » Voir d'autres dates

La profession de graphiste

http://perso.magic.fr

Pour réaliser les sites, une nouvelle profession émerge : celle de graphiste - graphic designer - dont les pages de présentation des CV témoignent.

Archive du 01 mai 2001 à 13:02 » Voir d'autres dates

L'héritage du Minitel

http://www.minitel.fr

L'accès aux services du Minitel par le Web se maintient durant les décennies 1990 et 2000. Le *form factor* des touches caractéristiques de la petite boîte beige est également repris.

Archive du 22 février 1997 à 03:56 » Voir d'autres dates

Le public handicapé du Web

http://www.starlingweb.com

Assez tôt la thématique de l'accessibilité du Web se pose. Les administrations et les entreprises promeuvent des conditions de navigation qui permettent à tous de consulter les ressources de la Toile mondiale.

Archive du 30 mai 1998 à 11:45 » Voir d'autres dates

Un monde en formulaire

http://www.cmtra.org

Aux côtés des liens hypertextuels, les formulaires se font une place de plus en plus importante sur le Web. Permettant la récolte d'informations sur les utilisateurs, ils sont l'interface privilégiée de la VPC.

Archive du 30 mai 1997 à 19:29 » Voir d'autres dates

Dites-le en GIF

http://www.tatoo.tm.fr

Avant de devenir le support-roi du détournement comique par les utilisateurs, le GIF est un format qui introduit le mouvement dans le monde de l'image fixe du Web des premiers temps. Léger et facile à réaliser, il colonise les pages dès 1996.

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Web Archives for Historians

Historians Who Use, Think About, and Work with Web Archives

Home | Please contribute! | Are you interested in web archives? | The Web Archive Bibliography | Who Are We?

The Historian of the Web : Crawler, Browser or Lurker?

March 13, 2015 | Uncategorized | 1990s, digital heritage, Louise Merzeau, Nineties, Web90 | peterwebster

[A special guest post by **Valérie Schafer** and **Francesca Musiani**.]

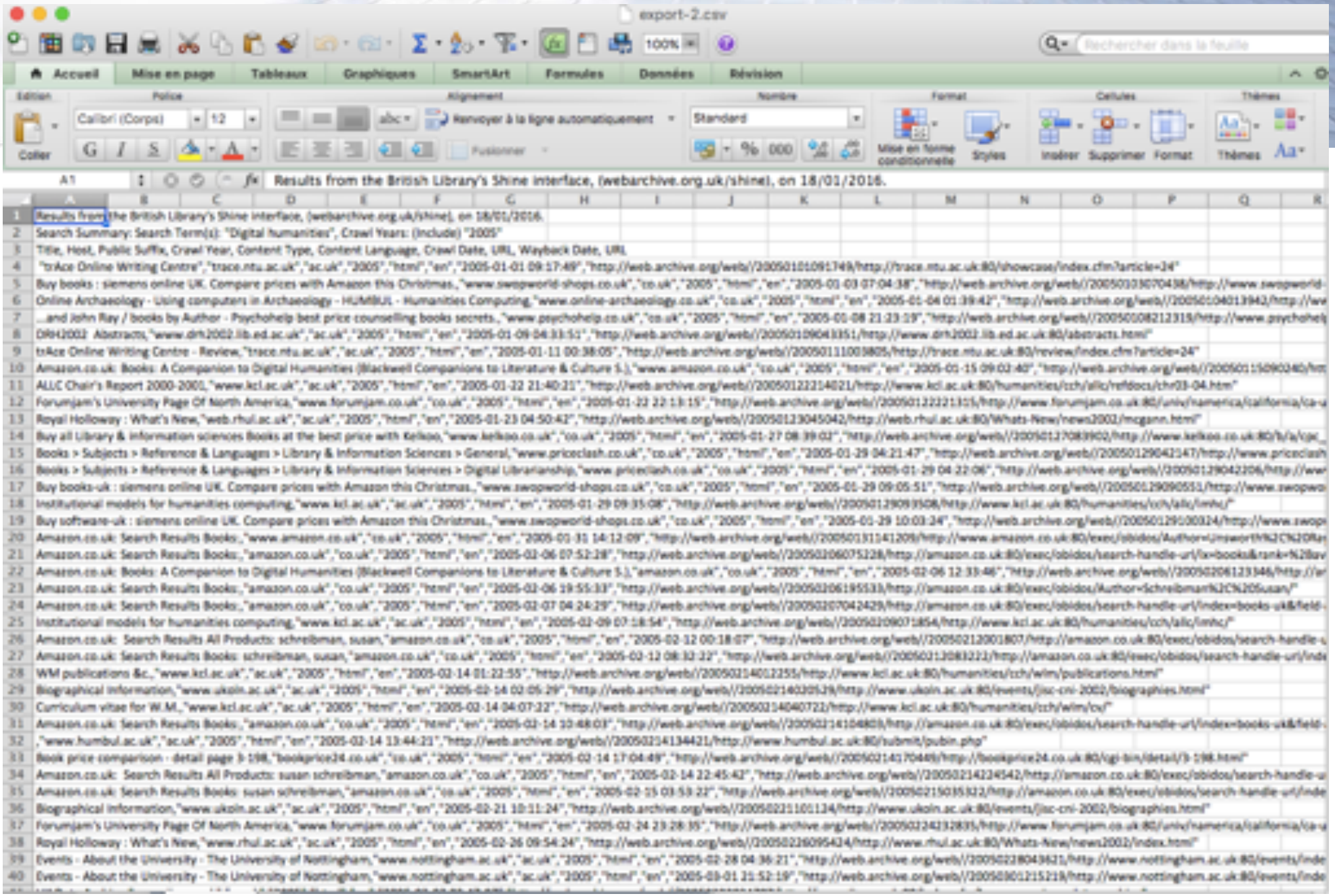
"They program us, we re-program them. They segment us, we move around. They accelerate, we linger. We can always be smarter than our machines."
[Louise Merzeau, « Le Flâneur Impatient », *Médium, Rythmes*, n°41, 2014/4, p. 20-29]

https://webarchivehistorians.org/2015/03/13/
the-historian-of-the-web-crawler-browser-or-lurker/

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY



Metadonnées dans SHINE

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

BnF Archives de l'internet Labs

COLLECTIONS MON COMPTE Aide A propos

En urgence, collecte sur les attentats parisiens de 2015

Statistiques

Des statistiques détaillées décrivent l'ensemble de la collection. Elles sont fournies sous forme de chiffres (nombre d'URL, de domaines, de formats, poids...) et de listes (répartition des URL par extension, 200 plus gros domaines...) et permettent d'appréhender la collection dans sa totalité.

Nom du fichier	Description	Date de mise à jour	Poids en octets	Télécharger
indicateurs-generaux-arc.txt	Principales statistiques sur le nombre d'URL ou de domaines, sur le poids, sur les formats...	16/06/2016	557	télécharger
tld-domain.txt	Répartition des 27 138 domaines collectés dans les 271 extensions (TLD)	16/06/2016	1 576	télécharger
tld-url.txt	Répartition des 14 784 657 URL collectées dans les 271 extensions (TLD)	16/06/2016	2 007	télécharger
domaine-url-top200.txt	Nombre d'URL collectées pour les 200 plus grands domaines	16/06/2016	3 707	télécharger
domaine-url_fr-top200.txt	Nombre d'URL collectées pour les 200 plus grands domaines en .fr	16/06/2016	3 602	télécharger
hote-url-top200.txt	Nombre d'URL collectées pour les 200 plus grands hôtes	16/06/2016	4 477	télécharger
annee-url.txt	Répartition des 14 784 657 URL collectées par année (2015)	16/06/2016	28	télécharger
codehttp-url.txt	Répartition des 14 784 657 URL collectées dans les 60 codes HTTP	16/06/2016	470	télécharger
typemime-url.txt	Répartition des 14 784 657 URL collectées dans les 3 262 formats (type MIME)	16/06/2016	106 261	télécharger

© BnF 2016 | Informations légales | allabs 1.2.0 (pfx)

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

(BnF

CHERCHER

Entre le et le Options avancées

#Hashtags fakenews

@Mentions Filtrer par mentions

Utilisateurs Filtrer par utilisateurs

Langages Filtrer par langage ex: fr,en

Retweets
- Chercher dans tous les tweets
- Exclure les tweets retweetés

Citations
- Chercher dans tous les tweets
- Exclure les tweets cités

Zoom Jour Semaine Mois Année Totalité

Recherche dans cette période De Mar 4, 2017 à May 11, 2017

50 523 résultats parmi les tweets archivés à toutes les dates pour le(s) hashtag(s) : **fakenews** dans tous les tweets dans toutes les citations (0.092 secondes)

©Ina

ina THEQUE

NAVIGATION

URL

Sites

RECHERCHE

Web

Vidéos

Twitter (beta)

STATISTIQUES

Chiffres clés

Feuille

AIDE

FAQ

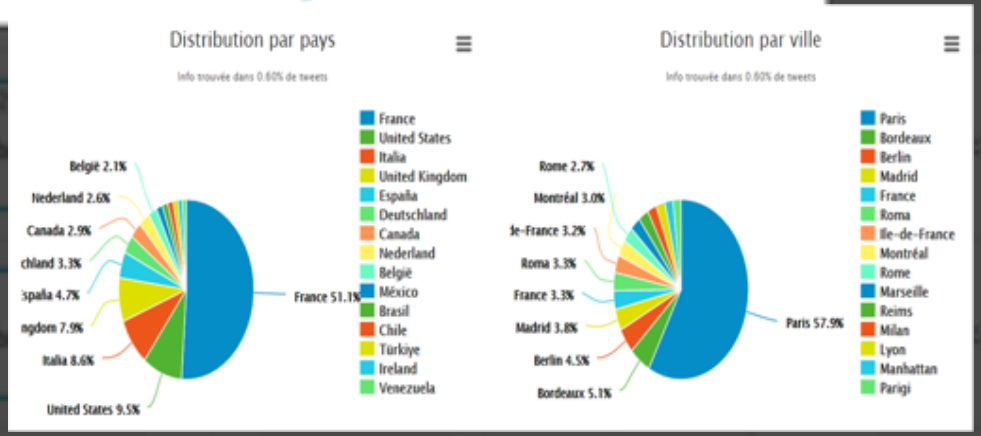
Liste | Analyse | Dashboard | Nuage de mots

Afficher 10 tweets par page

Afficher / Masquer les colonnes

Tweet	Utilisateur nom	Id	Date d'archivage
bibata sara @bibatasas • 2015-01-07 17:02:05 Mes pensés a la famille de se policiers abattu qui s'appelais Ahmed et musulman je suppose #CharlieHebdo	bibata sara	552857714775580672	2016-03-17 03:33:09
Isia c le sang @quentinmgn donc si j'ai bien compris ils defendent l' #JesuisCharlie #Je			6-03-17 10:34:37
Jayjay @JasonMalivernay RT @quentinmgn: donc si j'ai bien compris musulman? #JesuisChar			6-03-17 04:33:30
@bstoolAzra RT @quentinmgn: donc si j'ai bien compris musulman? #JesuisChar			6-03-17 04:34:46
To Neverland @5SecondsOfLumos • 2015-01-07 17:58:42 RT @quentinmgn: donc si j'ai bien compris ils defendent l'islam en al musulman? #JesuisCharlie #JesuisAhmed			56
Kenza @KenzaEl • 2015-01-07 18:07:38 RT @quentinmgn: donc si j'ai bien compris ils defendent l'islam en al musulman? #JesuisCharlie #JesuisAhmed			24
Laudinosaurus @Laudine_Tonks • 2015-01-07 18:45:29			

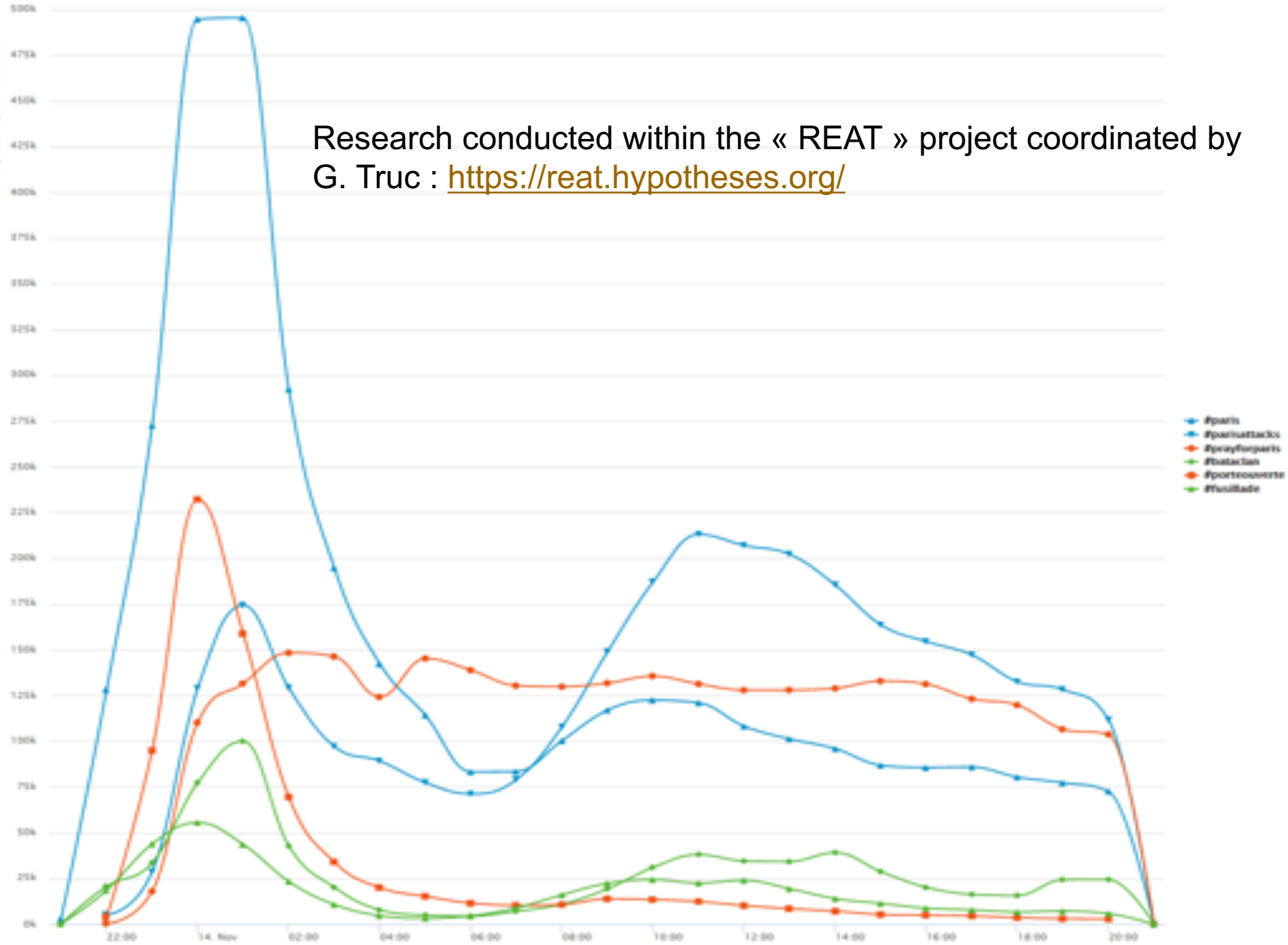
Tweet	Mentions	Utilisateur: date d'inscription
Date de tweet	Hôte	Id
Texte	Utilisateur	Lang
Nb de favoris	Utilisateur nom	Pays
Retweeté	Utilisateur: nb de followers	Location
Quoté	Utilisateur: nb de statuses	Date d'archivage
Nb de retweet	Utilisateur: location	Restore visibility
Tags	Utilisateur: lang	



©Ina

test ether 4350

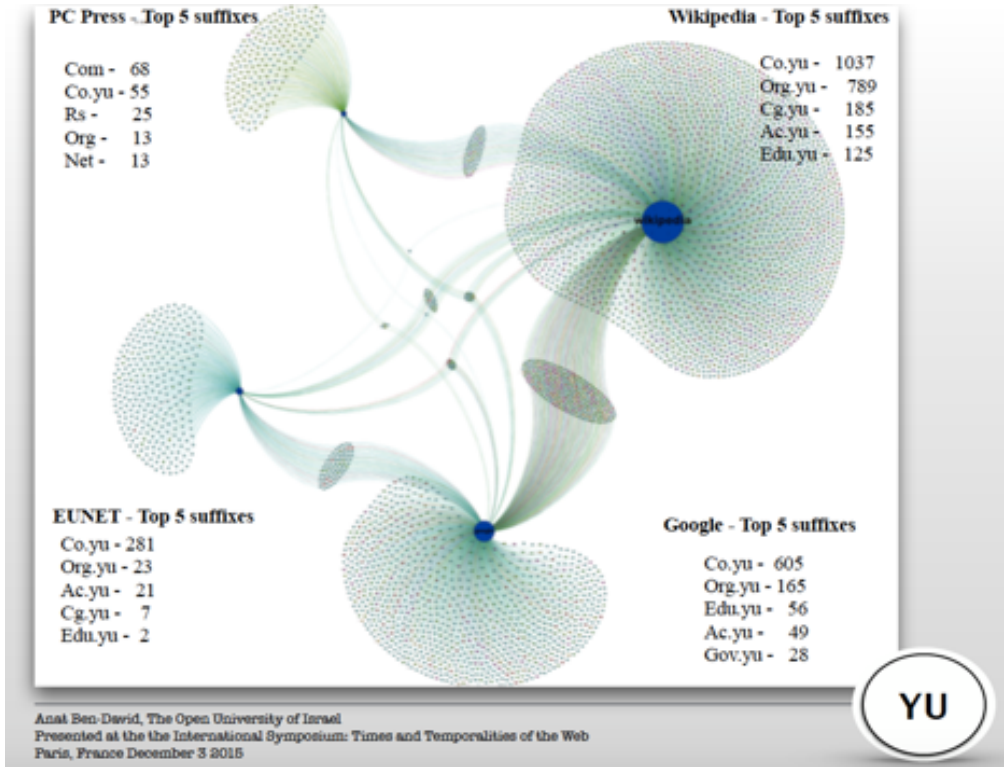
Research conducted within the « REAT » project coordinated by G. Truc : https://reat.hypotheses.org/



BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY



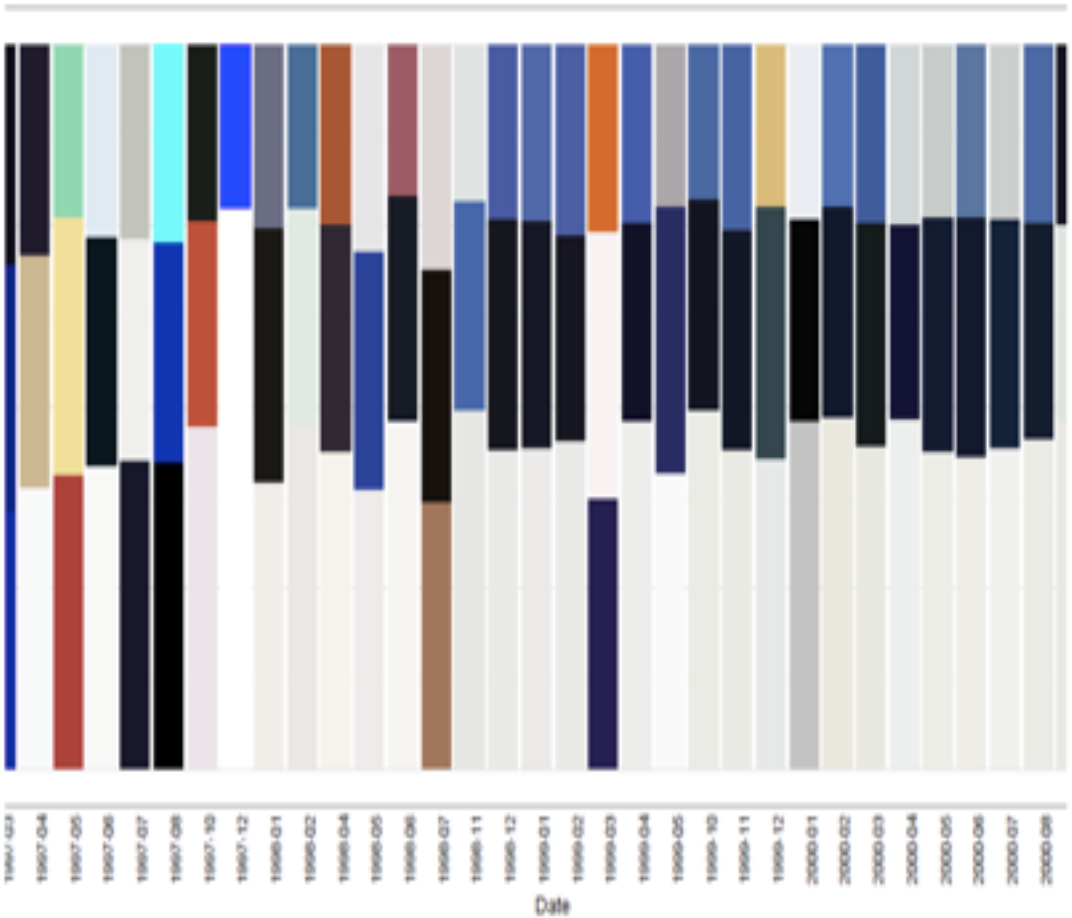
Ben-David, A. (2016a). What does the Web Remember of its Deleted Past? An Archival Reconstruction of the Former Yugoslav Top Level Domain. *New Media & Society*, 18(7), 1103–1119.

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

Distributions of colors in the Yugoslav Web



Ben-David, A., Amram, A., & R. Bekkerman. (2016).
The colors of the national Web: visual data analysis of the historical Yugoslav Web domain.
International Journal on Digital Libraries, DOI: 10.0007/s00799-016-0202-6.

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

EXPLORING THE GEOCITIES WEB ARCHIVE WITH WARCBASE & SPARK: LINKS (OR HOW WE CAN USE WARCBASE TO FIND AMAZING SITES TO ASK HISTORICAL QUESTIONS!)

23 February 2016

Not just spaghetti and meatballs, but the starting point for research.

In my last post, we left off with scripts running to extract all URLs and a link diagram. They finished decently quickly – about three days on our rho server at York University, or about 30 minutes on our roaringly-fast cluster. Given that hopefully we will be running these only once or twice at first, even three days isn't too bad.

This is a relatively technology-heavy post, but I think even relative newcomers to digital history might find this interesting as an entree into several tools that we might use to extract meaningful historical information from big datasets.

https://ianmilligan.ca/2016/02/23/exploring-the-geocities-web-archive-with-warcbase-spark-links-or-how-we-can-use-warcbase-to-find-amazing-sites-to-ask-historical-questions/

BnF

UNIVERSITÉ DU LUXEMBOURG

C²DH LUXEMBOURG CENTRE FOR CONTEMPORARY AND DIGITAL HISTORY

“The question is not *if* scholarly studies within the Humanities want to “go digital”, but rather *how*”. (p. 9)

“Although Moretti in the main uses the distant reading approach to the study of large amounts of digital data, I will argue that none of these two approaches are per se inscribed or inherent in the digital material. By this I mean that simply because collections of digital material are in many cases big data, which opens the possibility of asking and answering new types of research questions, this does not necessarily mean that they have to be approached as Big Data”. (p. 11)

Brügger N., Humanities, *Digital Humanities, Media Studies, Internet Studies: An inaugural Lecture,* Aarhus, CFI, 2015.

Merci de votre attention !

Pour plus d'informations, voir les pages dédiées sur le site de la BnF :

- Aspects techniques et préservation : http://www.bnf.fr/fr/professionnels/innovation_numerique.html
- Collections et parcours guidés : http://www.bnf.fr/fr/collections_et_services/livre_presse_medias/a.archives_internet.html
- Bibliographie : http://www.bnf.fr/documents/bibliographie_dl_web.pdf
- Procédure du dépôt légal : http://www.bnf.fr/fr/professionnels/depot_legal/a.dl_sites_web_mod.html

- Twitter : @DLWebBnF

- Le site du projet Web90: https://web90.hypotheses.org

BnF