



{ BnF

## IFLA International Newspaper Conference

**“Newspaper Digitization and Preservation.  
New prospects.  
Stakeholders, Practices, Users and Business Models”**

**11-13 April 2012  
BnF, Paris**

With the support of:



**Isako**

Bookkeeper

EUROPRESSE.COM  
an initiative of CEDROMISV

**PLANMAN  
TECHNOLOGIES**



**diadeiS**  
groupe numeris

{BnF



# Legal Deposit of Online Newspapers

*Digital collections in BnF stacks*

Clément Oury

Head of Digital Legal Deposit  
Bibliothèque nationale de France





## Summary



- The issue : ensuring the continuity of BnF heritage collections
- Legal and technical solutions
- Insight within BnF press collections
- Challenges and new projects



## Ensuring the continuity of BnF newspaper collections

- Through legal deposit, BnF has been collecting all major newspaper titles since... the invention of periodicals
- This mission is now faced with challenges due to development of online press
  - Digital migration of paper publications (“bi-media” or digital only)
  - Growing role of “pure-player”
  - As a paradox, increasing number of paper editions



## Avoiding the “digital memory gap”

- This issue is not limited to the press
  - All kinds of heritage material are undergoing a digitization process : books, images, sounds, videos...
- Heritage institutions such as BnF need to find the legal and technical means to tackle these issues
- On the legal side, the solution has been found in BnF’s long-standing mission: legal deposit



## The legal deposit framework



- Legal deposit : since 1537, each editor should send copies of their production to the ~~royal imperial~~ national library
- Legal deposit has evolved over time to cover different media types (printed books, engravings, now DVDs, software...)
- 2006: legal deposit extended to “signs, signals, writings, images, sounds or messages of any kind communicated to the public by electronic means”
- This mission is shared with the National Audiovisual Institute (INA) for radio and television websites
- The goal is not to gather the « best of the web », but to preserve a collection representative of the web at a certain date





## A matter of harvesting



- Using a software called “robot”, “spider” or “harvester”, which
  - Departs from a list of “seeds” URLs
  - Extracts hyperlinks from web pages and follows them... just like an automated internet user
  - Copies only pages and files that are in its scope (defined by curators)
- From a technical point of view, it is not a “deposit” anymore but a collect

# {BnF



# The true face of the robot



Status as of Feb. 26, 2007 17:39:01 GMT Alerts: no alerts

HOLDING JOBS

[View logs](#)

0 jobs pending, 97 completed

[Console](#) [Jobs](#) [Profiles](#) [Logs](#) [Reports](#) [Setup](#) [Help](#)

View: [crawl.log](#)

[local-errors.log](#)

[progress-statistics.log](#)

[runtime-errors.log](#)

[uri-errors.log](#)

By: [Line number](#)

[Time stamp](#)

[Regular expression](#)

[Tail](#)

Line number:

Lines to show:

crawl.log for elec2007-20070205124900-f2.host4.1

```
2007-02-07T10:02:54.3002 200 3300 http://dsk.typepad.com/dsk.JPG EXEE http://dsk.typepad.com/
2007-02-07T10:02:54.7122 200 35638 http://assurancebayrou.hautetfort.com/archive/2007/02/04/je
2007-02-07T10:02:54.7162 200 58 http://changerlarepublique.over-blog.com/sms-1170849842.xml
2007-02-07T10:02:54.9532 200 42729 http://dejademain.hautetfort.com/archive/2007/02/04/index.ht
2007-02-07T10:02:54.9582 200 43096 http://aiglon78.noosblog.fr/le_blog_daiglon/atom.xml E http:
2007-02-07T10:02:54.9632 200 906 http://pics.lesechos.fr/images/elections-2007/autres_cand_en
2007-02-07T10:02:54.9762 200 474 http://www.jeunes-udf.org/include/orange.css PRE http://www.jeunes-udf.org/
2007-02-07T10:02:54.9782 200 4268 http://stats.wordpress.com/w.js?4 EXEE http://elections2007.wordpress.com/2007/02/06/63-dindecis/
2007-02-07T10:02:54.9822 1 69 dns:gabrielbisiaux.hautetfort.com EXEP http://gabrielbisiaux
2007-02-07T10:02:54.9862 200 92792 http://2007.blog.20minutes.fr/index.rss LE http://2007.blog.
2007-02-07T10:02:55.1222 200 39797 http://christianlajon.club-blog.fr/weblog_de_christian_lajon
2007-02-07T10:02:55.1232 200 2905 http://jcsfrance.free.fr/spip_style.css E http://jcsfrance.f
2007-02-07T10:02:55.1262 200 4826 http://synthesenationale.hautetfort.com/images/thumb_pinoche
2007-02-07T10:02:55.1272 404 37 http://cambadelis.over-blog.net/Mozilla/3 X http://cambadeli
2007-02-07T10:02:55.1412 200 103784 http://jppf29.free.fr/graphiques/jeunes_assemblee_nationale_1
2007-02-07T10:02:55.3362 200 7263 http://pub.mybloglog.com/comm2.php?mb1ID=20060914061512816c_
2007-02-07T10:02:55.3382 200 50298 http://www.generation2007.fr/?2006/10/04/82-les-candidats-zo
2007-02-07T10:02:55.3412 404 368 http://image.lycos.fr/xml/reuters/2006-11-11T160150Z_01_NOOT
2007-02-07T10:02:55.3442 200 5067 http://jean-micheljardry.hautetfort.com/images/thumb_Peter_M
2007-02-07T10:02:55.3482 1 66 dns:notzugzwang.blogspot.com EXEP http://notzugzwang.blogsp
2007-02-07T10:02:55.3512 1 63 dns:www.politique-info.org EXEP http://www.politique-info.or
2007-02-07T10:02:55.3602 200 1893 http://www.celeonet.fr/404.php EPXPR http://www.bigbangblog.
2007-02-07T10:02:55.4572 200 59938 http://www.milletelechargements.com/?id=kochon&w=9 EPRE http
2007-02-07T10:02:55.4692 200 18607 http://christan-garino-esperanto.over-blog.org/article-53961
2007-02-07T10:02:55.5662 200 1535 http://acguingamp.blog.mongenie.com/doc/blog/blog.css E http
2007-02-07T10:02:55.8712 200 47475 http://www.kino.de/newinfo.php4 EEELPR http://www.kino.de/rc
2007-02-07T10:02:55.8732 1 62 dns:www.blogueparade.com EP http://www.blogueparade.com/imag
2007-02-07T10:02:55.8762 1 73 dns:tkfiles.storage.msn.com EP http://tkfiles.storage.msn.com/xrpijpk0x-0gn-0unbrkxvz0n1f0v4r000q0-0lpin0Xv5sY4JyI
2007-02-07T10:02:55.8772 1 94 dns:sc2.sclive.net EP http://sc2.sclive.net/00.00.0000/Web/images/c.gif text/dns #065 20070207100254108+171 3
2007-02-07T10:02:55.8802 1 94 dns:sc4.sclive.net EP http://sc4.sclive.net/00.00.0000.0000/Web/images/arrow_rightg.gif text/dns #078 200702071002
2007-02-07T10:02:55.8812 404 2515 http://dordogneavecmam.blogup.fr/www.blogup.com/rg-simple-format X http://dordogneavecmam.blogup.fr/ text/html #12
2007-02-07T10:02:55.8842 200 191 http://blog.20minutes.fr/images/extras/atom10.gif E http://lesmousquetairesdesegolene.blog.20minutes.fr/ image/gif
2007-02-07T10:02:55.8902 404 317 http://red.ads.t-online.de/ EX http://a.as-eu.falkag.net/dat/dlv/aslmain.js text/html #023 20070207100254122+47 HF
2007-02-07T10:02:55.8932 1 97 dns:shared.live.com XP http://shared.live.com/5kYRRmG3cfY1NTwrK8iT4zJ1c7SEuQvV5xvzPKAhVYg-VbTmOZREGHiAj9fE1lau6Y-
2007-02-07T10:02:55.8962 1 55 dns:help.live.com EP http://help.live.com/resources/neutral/launchhelp.js text/dns #007 20070207100253792+73 3I42F
2007-02-07T10:02:55.8972 200 43 http://12.sedo.com/log/n.gif?sa=16779&gf=0&cs=572000000&x=3840&y=577&id=esi&si=1&ze=1170842569 EXE http://www.techne
```



```
34NJI6EGOI
fort.com/
text/xml #
html #035
#128 2007
age/jpeg #
application/x-je
-verts-sur
7 20070207
fr/ applic
BRGRFMRYDI
ag/mpf ime
52867+76 1
/jpeg #020
heading_t
632+129 2F
5/10/04/83
20061111.
etfort.com
D 3I42H3S6
7100252339
632+129 2F
2007020710
ver-blog.c
31 2007020
B5QTRPVI
222+96 3I4

```



# {BnF



## If it does not work...



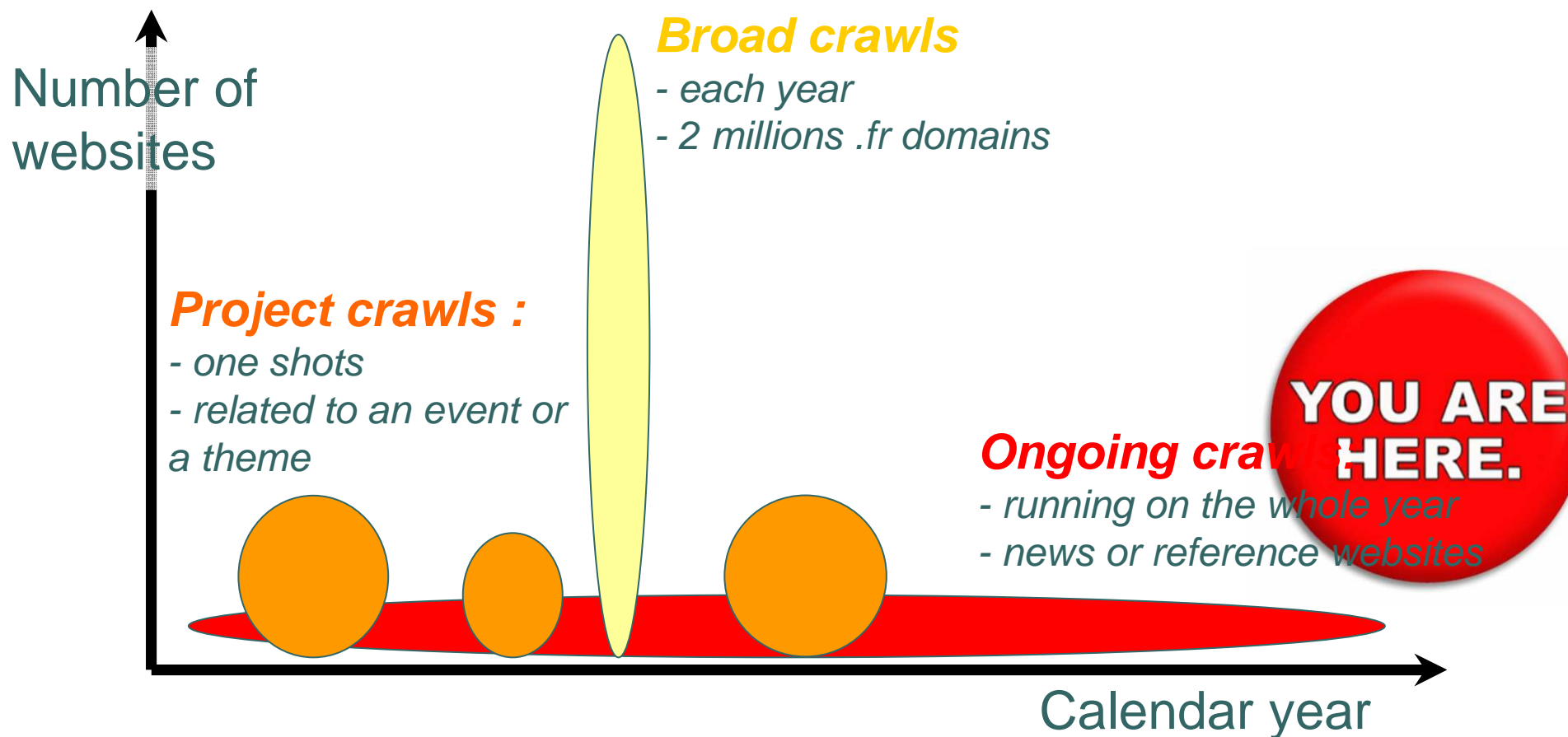
www.shutterstock.com · 56333926

- Robots may encounter technical issues and obstacles
  - Password-protected content
  - Subscription content
  - Complex technical architectures (flash, javascripts, etc...)
- The heritage code takes this case into account
  - Web sites editors shall help BnF harvesting their website by giving codes and passwords if needed
  - Deposit may be used if automatic harvesting is not feasible

{BnF



# BnF “mixed model” of harvesting



{BnF



# Harvesting the “news” at BnF

- **100 titles** selected and curated by
  - The press service (Law, Economics, Politics Department)
  - The periodicals service (Legal Deposit Department)
- According to a **typology** defined by the press service
  - Press agencies
  - National daily newspapers
  - Regional daily newspapers
  - Magazines
  - Portals
  - Internet information
  - Pure players

Affiner

Affiner par

Thème

- agence (1) >
- infonet (4) >
- portail (9) >
- presse magazine (8) >
- presse quotidienne nationale (13) >
- presse quotidienne régionale (47) >
- pure player (15) >

État

- actif (95) >
- inactif (2) >

Type de collecte

- ciblée (97)

Fréquence

- 1 fois par jour (97)

Collecte Actualités (ACTU)

Résultats : 97

Vue experte

Fiche	URL de départ ▼	État	Type de collecte	Fréquence
	<a href="http://actu.orange.fr">http://actu.orange.fr</a>	Actif	Ciblée	1 fois par jour
	<a href="http://fr.news.yahoo.com">http://fr.news.yahoo.com</a>	Actif	Ciblée	1 fois par jour
	<a href="http://fr.reuters.com">http://fr.reuters.com</a>	Actif	Ciblée	1 fois par jour
	<a href="http://infopolitique.free.fr">http://infopolitique.free.fr</a>	Actif	Ciblée	1 fois par jour
	<a href="http://news.fr.msn.com/m6-actualite/">http://news.fr.msn.com/m6-actualite/</a>	Actif	Ciblée	1 fois par jour
	<a href="http://news.google.fr">http://news.google.fr</a>	Actif	Ciblée	1 fois par jour
	<a href="http://owni.fr">http://owni.fr</a>	Actif	Ciblée	1 fois par jour
	<a href="http://quoi.info">http://quoi.info</a>	Actif	Ciblée	1 fois par jour
	<a href="http://tempsreel.nouvelobs.com">http://tempsreel.nouvelobs.com</a>	Actif	Ciblée	1 fois par jour
	<a href="http://www.20minutes.fr">http://www.20minutes.fr</a>	Actif	Ciblée	1 fois par jour



## Access to web archives

- Onsite access only (for IP rights and data protection act enforcement)
- Access is restricted to “researchers”
  - Not only scholars, but all citizens that have a demonstrated need to access web archives
- Access is provided on all computers in all BnF research reading rooms
  
- Access will be opened in main regional libraries

{BnF



# Insight into BnF collections

How to access web archives



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.

### Le Web vert : les politiques du développement durable

Entreprises, pouvoirs publics, militants ou citoyens : ces acteurs ont fait du développement durable un sujet majeur de préoccupation et de communication. Le Web est le témoin privilégié de ces dynamiques.



### Le web militant

Internet est un canal de diffusion d'informations, de mots d'ordre, un lieu de débat, de participation et d'action. Ce dossier illustre la diversité de ces usages militants.



# Browsing the archives of the online newspapers

**Avertissement :** les techniques utilisées ne permettent pas l'archivage de tous les sites, ni la conformité des archives aux sites originaux.

BnF - Archives de l'Internet - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://archivesinternet.bnf.fr/

BnF - Archives de l'Internet

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...

**Avertissement:** les techniques utilisées...

## 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.


### Le Web vert : les politiques du développement durable

Entreprises, pouvoirs publics, militants ou citoyens : ces acteurs ont fait du développement durable un sujet majeur de préoccupation et de communication. Le Web est le témoin privilégié de ces dynamiques.



### Le web militant

Internet est un canal de diffusion d'informations, de mots d'ordre, un lieu de débat, de participation et d'action. Ce dossier illustre la diversité de ces usages militants.



Terminé



BnF - Archives de l'Internet - recherche par url - résultats - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://archivesinternet.bnf.fr/query?type=urlquery&url=http%3A%2F%2Fwww.lemonde.fr&uiUrl=http%3A%2F%2Fwww.lemonde.fr&Submit.x=91&Submit.y=8&S

BnF - Archives de l'Internet - re...

Retrouver ce site ou cette page :  
http://www.lemonde.fr  
**Remonter le temps**

**Option**  
Limiter à : [ ] [ ] [ ]

**Recherche par URL** 1 084 résultats  
pour : http://www.lemonde.fr du 1 jan. 1996 au 30 juin 2011  
[ouvrir tout](#) | [fermer tout](#)

2011		303 résultats ▶
2010		369 résultats ▶
2009		59 résultats ▶
2008		67 résultats ▶
2007		206 résultats ▶
2006		13 résultats ▶
2005		10 résultats ▶
2004		2 résultats ▶
2003		10 résultats ▶
2002		2 résultats ▶
2001		36 résultats ▶
2000		6 résultats ▶

Terminé

Retrouver ce site ou cette page :

http://www.lemonde.fr

Remonter le temps

Option

limiter à :

Three dropdown menus for search options.



### Recherche par URL

1 084 résultats

pour : http://www.lemonde.fr

du 1 jan. 1996 au 30 juin 2011

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

2011  303 résultats ▶

2010  369 résultats ▶

2009  59 résultats ▶

2008  67 résultats ▶

2007  206 résultats ▼

jan.	fév.	mar.	avr.	mai	juin	jui.	août	sep.	oct.	nov.	déc.
2 jan.	5 fév.	5 mar.	2 avr.	2 mai	4 juin	2 jui.	27 août	20 sep.	3 oct.	7 nov.	
2 jan.	5 fév.	5 mar.	2 avr.	2 mai	4 juin	2 jui.	27 août		12 oct.	7 nov.	
3 jan.	5 fév.	5 mar.	2 avr.	2 mai	7 juin	2 jui.	27 août		12 oct.	7 nov.	
3 jan.	6 fév.	5 mar.	2 avr.	2 mai	7 juin	4 jui.	27 août			7 nov.	
3 jan.	7 fév.	6 mar.	2 avr.	2 mai	7 juin	4 jui.				7 nov.	
3 jan.	7 fév.	6 mar.	2 avr.	2 mai	7 juin	4 jui.				8 nov.	
4 jan.	8 fév.	8 mar.	2 avr.	2 mai	8 juin	4 jui.				11 nov.	
9 jan.	8 fév.	8 mar.	3 avr.	3 mai	11 juin	4 jui.				26 nov.	
9 jan.	8 fév.	8 mar.	3 avr.	3 mai	11 juin	4 jui.				26 nov.	
9 jan.	12 fév.	12 mar.	4 avr.	3 mai	11 juin	5 jui.					
15 jan.	12 fév.	12 mar.	10 avr.	5 mai	11 juin	5 jui.					
15 jan.	12 fév.	12 mar.	10 avr.	5 mai	11 juin	9 jui.					
15 jan.	12 fév.	12 mar.	10 avr.	6 mai	11 juin	9 jui.					
22 jan.	14 fév.	19 mar.	10 avr.	7 mai	12 juin	10 jui.					
22 jan.	19 fév.	19 mar.	16 avr.	7 mai	12 juin	10 jui.					
22 jan.	19 fév.	19 mar.	16 avr.	7 mai	12 juin	10 jui.					
24 jan.	19 fév.	19 mar.	16 avr.	7 mai	12 juin	10 jui.					
31 jan.	19 fév.	26 mar.	16 avr.	7 mai	13 juin	10 jui.					
31 jan.	19 fév.	26 mar.	16 avr.	7 mai	13 juin	10 jui.					
31 jan.	21 fév.	26 mar.	17 avr.	7 mai	13 juin						
	26 fév.	27 mar.	17 avr.	9 mai	18 juin						
	26 fév.		17 avr.	9 mai	18 juin						
	26 fév.		21 avr.	9 mai	18 juin						

Le Monde.fr : A la une - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://www.lemonde.fr/

Le Monde.fr : A la une

http://www.lemonde.fr/ OK Version 171 sur 939 - 06 mai 2007 à 16:34

Archives de l'Internet  
Accès expérimental  
accueil

première précédente suivante dernière

Recherchez depuis 1 mois


sur Le Monde.fr sur le web avec YAHOO! SEARCH

Identifiez-vous  
email  
Mot de passe  
\*\*\*  
 mémorisez | oublié?

Recevez les newsletters gratuites Ajoutez Le Monde.fr à vos favoris

actualités perspectives pratique annonces le desk le kiosque newsletters multimédia références EDITION ABONNES Abonnez-vous 6€ Jeudi 30 juin 2011

## Le taux de participation s'élève à 75,11 % à 17 heures en France métropolitaine



Ce taux est le plus élevé jamais enregistré à cette heure depuis 1965, selon le ministère de l'intérieur. Selon des projections des instituts de sondage, il serait de 85 %, selon l'IFOP ou de 86 %, selon CSA et Ipsos.

**Heure par heure** Très forte mobilisation des électeurs, lors du second tour

**Les faits** Une forte mobilisation policière prévue pour le second tour de la présidentielle

**Portfolio** "Il n'y a pas de suspense"

**Portfolio** A Belleville, les électeurs de François Bayrou à l'heure du choix

**Vous** Vos photos du second tour de la présidentielle

**Infographie** Taux de participation à la mi-journée pour le second tour de l'élection présidentielle depuis 1969

**Vidéos** La journée de vote du second tour

**LES DÉPÊCHES**

Tennis: la finale du tournoi WTA de Varsovie reportée Reuters 17h58

Serie A: Ascoli et Messine relégués, le Milan AC quatrième Reuters 17h56

La participation totale serait de 85% à 86%, selon Ifop et CSA Reuters 17h46

Cyclisme: Thomas Dekker remporte le Tour de Romandie Reuters 17h41

► Toutes les dépêches

**VOS RÉACTIONS**

"les experts sont au contraires unanimes: le vote électronique n'est en rien vérifiable, et présente des risques élevés de fraude et de truquages en tout genre. Et ce n'est pas un texte datant de 1969 (!), préhistoire de l'informatique, qui peut rendre crédible l'autorisation du ministère de l'intérieur, pas plus que les certifications bidons ne répondant à aucun standard sérieux de sécurité (pour rappél. niveau faible

**PRÉSIDENTIELLE 2007: HEURE PAR HEURE** Toute l'édition spéciale

18h11 **Portfolio** "Il n'y a pas de suspense"

17h05 **Portfolio** A Belleville, les électeurs de François Bayrou à l'heure du choix

16h39 **Infographie** Taux de participation à la mi-journée pour le second tour de l'élection présidentielle depuis 1969

16h15 **Vous** Vos photos du second tour de la présidentielle

15h17 **Vidéo** Vote électronique à Issy-les-Moulineaux

15h06 | Réactions (8)

**Les faits** Une forte mobilisation policière prévue pour le second tour de la présidentielle

Terminé

Le Monde.fr : A la une - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://www.lemonde.fr/

Le Monde.fr : A la une

http://www.lemonde.fr/ OK Version 172 sur 939 : 07 mai 2007 à 09:48

Archives de l'Internet Accès expérimental accueil

première précédente suivante dernière

Recherchez depuis 1 mois


sur Le Monde.fr sur le web avec YAHOO! SEARCH

Recevez les newsletters gratuites Ajoutez Le Monde.fr à vos favoris

Identifiez-vous email Mot de passe \*\*\* mémorisez oublié?

actualités perspectives pratique annonces le desk le kiosque newsletters multimédia références EDITION ABONNÉS Abonnez-vous 6€ Jeudi 30 juin 2011

## Les socialistes tentent de dépasser les conflits de personnes pour débattre du fond



"Mon problème n'est pas un problème de personne, en aucune manière", précise Dominique Strauss-Kahn, lundi, sur Europe 1. Il ne s'agit pas de "droitiser la gauche", met en garde Laurent Fabius. François Hollande réclame l'unité à l'approche des législatives.

**Les faits** Ségolène Royal lance l'offensive pour s'imposer comme leader naturel du PS  
**Analyse** La solitude volontaire et subie de Mme Royal explique son échec  
**Son** Jean-Marie Colombani : "Nicolas Sarkozy a renoué l'idéologie de sa famille politique"  
**Télézapping** La soirée présidentielle à la télévision  
**Heure par heure** La France se réveille avec un nouveau président  
**Revue de Web** Les échauffourées de la nuit du 6 au 7 mai sur Dailymotion  
**Résultats** Les résultats du second tour par département

**PRÉSIDENTIELLE 2007: HEURE PAR HEURE** Toute l'édition spéciale

11h16 **Revue de Web** Place de la Concorde, 30 000 personnes fêtent la victoire de M. Sarkozy

10h56 **Son** Philippe Mesmer : Au Japon, "des réactions plutôt limitées"

11h03 **Réactions** Les socialistes tentent de dépasser les conflits de personnes pour débattre du fond

10h53 **Bilan** La gauche a perdu la région Nord et manqué de voix bayrouistes

**PRÉSIDENTIELLE LES RÉSULTATS 2d tour**  
 Les résultats par ville et par département.

Votre département  
 Votre choix

Carte des résultats

**LES DÉPÊCHES**

Aucun rescapé dans l'accident du Boeing kenyan au Cameroun Reuters 11h24

Aucun rescapé dans l'accident du Boeing kenyan au Cameroun Reuters 11h24

DSK lie la défaite de Ségolène Royal au manque de clarté du PS Reuters 11h22

Ultimatum à nouveau repoussé pour l'otage français des taliban Reuters 11h04

Toutes les dépêches

**VOS RÉACTIONS**

"Enfin la France va sortir de

Terminé

Le Monde.fr : A la une - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://www.lemonde.fr/

Le Monde.fr : A la une

http://www.lemonde.fr/ OK Version 174 sur 939 07 mai 2007 à 12:09 Archives de l'Internet Accès expérimental

première précédente suivante dernière

Recherchez depuis 1 mois

sur Le Monde.fr sur le web avec Yahoo! SEARCH


Recevez les newsletters gratuites Ajoutez Le Monde.fr à vos favoris

Identifiez-vous email Mot de passe \*\*\* mémorisez oublié?

Le Monde.fr Mise à jour à 13 h 54 - Paris

actualités perspectives pratique annonces le desk le kiosque newsletters multimédia références EDITION ABONNES Abonnez-vous 6€ Jeudi 30 juin 2011

## Nicolas Sarkozy, la conquête méthodique du pouvoir



Il y songe depuis toujours, et sérieusement depuis au moins... vingt-cinq ans. Nicolas Sarkozy s'est préparé à la présidentielle en professionnel, après son échec aux européennes de 1999. Ces derniers jours, il se retournait encore avec orgueil sur son histoire : "Si je suis élu président, je l'aurai bien mérité."

**Chat à 14 h 30** L'analyse du score de Nicolas Sarkozy

**Les faits** Les prochains jours de Nicolas Sarkozy

**Heure par heure** Dès lundi matin, les législatives sont sur toutes les lèvres

**Editorial** Changer, par Jean-Marie Colombani

**Panorama** La deuxième tour et ses résultats vus par la télévision

**Vous** Vos explications à la défaite du PS : "Personnellement, j'ai voté socialiste, pas Royal"

**Revue de Web** Les échauffourées de la nuit du 6 au 7 mai sur Dailymotion

**PRÉSIDENTIELLE 2007: HEURE PAR HEURE** Toute l'édition spéciale

13h54 **Les faits** Les petits partis vont tenter de sauver leurs élus à l'Assemblée nationale

13h54 **Eclairage** La droite aborde le scrutin de juin en position de force

13h33 **Eclairage** Une forte participation qui dément la consigne frontiste

13h33 **Les faits** Jean-Marie Le Pen entend "marquer à la culotte" le nouveau président

13h13 **Les faits** Les organisations à la gauche du PS appellent à la "résistance"

13h13 **Reportage** De vivats en rappels, d'un balcon à l'autre, la candidate goûte, une dernière fois, le parfum de la gloire

**PRÉSIDENTIELLE LES RÉSULTATS 2d tour**

Votre commune

ou votre code postal

Votre département

Votre choix

Carte des résultats

**LES DÉPÊCHES**

"Restons unis" lance Ségolène Royal au lendemain de sa défaite Reuters 13h58

Nicolas Sarkozy en retraite dans le sud de la Corse ? Reuters 13h48

Deux Palestiniens tués dans des heurts au Sud-Liban Reuters 13h46

Un fonctionnaire du Vatican condamné pour détention de drogue Reuters 13h42

Toutes les dépêches

Terminé

{BnF



# Insight into BnF collections

Variety of harvested publications



National  
daily  
press

The screenshot shows the France Soir website interface. At the top, there is a navigation bar with the logo, a search bar, and utility links like 'S'identifier' and 'Recevoir la newsletter'. Below this is a secondary navigation bar with categories: ACTUALITÉS, PEOPLE / TV, SPORTS, PRATIQUE, LOISIRS, TURF, and PRÉSIDENTIELLE 2012. The main content area features a large article about Eden Hazard, titled 'Hazard "Le meilleur joueur du championnat"', with a photo of him celebrating. To the right is a 'DERNIÈRE MINUTE' section with a list of news items. Below the main article are four smaller thumbnail images with captions: 'Abstention : Hollande fait le point', 'Eva Joly : Elle sortira de l'hôpital ce lundi', 'Disneyland Paris : Mickey fête ses 20 ans', and 'Hazard, "le meilleur joueur du championnat"'. At the bottom left, there is a section for 'L'ESSENTIEL' with a sub-section for 'PRÉSIDENTIELLE 2012' featuring an article titled 'Exclusif - Rama Yade : "Mélenchon met en lumière la fadeur de Hollande"'. To the right of this is a 'PRÉSIDENTIELLE' section with a video player titled 'Le Pen vs. Lapix : round 2'. On the far right, there is a 'SONDAGE' section titled 'A 20 jours du premier tour de la présidentielle, êtes-vous certain(e) d'aller voter?' with radio button options and a 'Voter' button. The page footer includes a URL, version information, and a date stamp.

France Soir 100% NUMÉRIQUE - L'INFO CONTINUE .FR

Présidentielle 2012  
Euro 2012 football  
Hunger Games, le prochain Twilight ?

100% NUMÉRIQUE - L'INFO CONTINUE .FR

ACTUALITÉS PEOPLE / TV SPORTS PRATIQUE LOISIRS TURF PRÉSIDENTIELLE 2012

» Téléchargez l'application iPhone / iPad

## Hazard

"Le meilleur joueur du championnat"

**DERNIÈRE MINUTE**

- 09:59 Plateforme Total Elgin : Le plan d'intervention en préparation
- 09:59 Présidentielle : Mélenchon ne veut pas "d'émeute"
- 09:39 Créteil : "J'ai emmené ma fille de 4 ans à l'abattoir"
- 09:38 Johnny chante et répète à Los Angeles, la ville qui lui a "sauvé la vie" (afp)
- 09:10 Birmanie: Suu Kyi salue une "ère nouvelle" dans un discours de victoire (afp)

» Toute l'actualité en direct

### SONDAGE

**A 20 jours du premier tour de la présidentielle, êtes-vous certain(e) d'aller voter ?**

- Oui, aucune hésitation
- Oui mais je ne sais pas encore pour qui
- Non je me déciderai au dernier moment
- Non d'ailleurs, c'est sûr, je n'irai pas

Voter

702 votants

**L'ESSENTIEL**

**PRÉSIDENTIELLE 2012**

## Exclusif - Rama Yade : "Mélenchon met en lumière la fadeur de Hollande"

Après avoir, dit-elle, « pris le temps de la réflexion »,

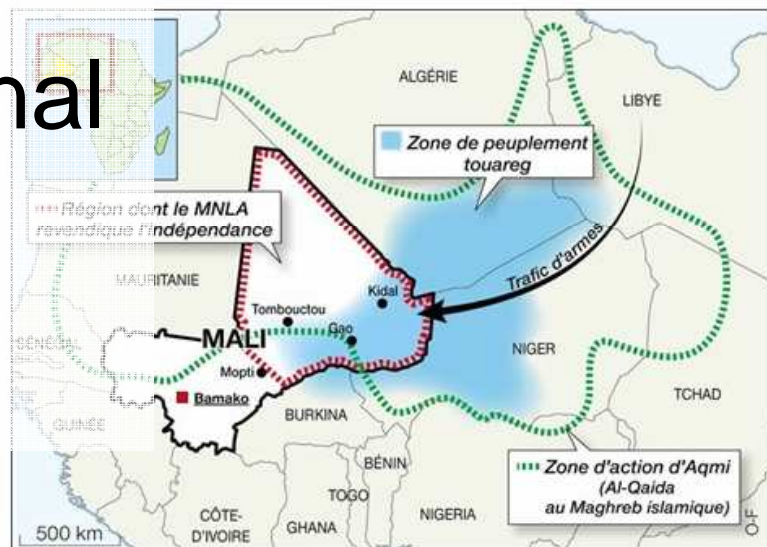
**PRÉSIDENTIELLE**

Le Pen vs. Lapix : round 2

http://www.francesoir.fr/sites/all/themes/ OK  
Version 50 sur 53 : 02 avril 2012 à 08:14  
Accès permanent accueil

# Mali. Le pays coupé en deux par les rebelles et les militaires

MALI conflits



# Regional daily press

### En direct de nos rédactions

- A la Une** 09:40 **Mali - Conflits**  
Mali. Le pays coupé en deux par les rebelles et les militaires
- Bretagne** 09:32 **France - Transports**  
Grève des contrôleurs aériens. Deux jours de perturbation
- Basse Normandie** 09:29 **France - Économie**  
Automobile. Le marché des voitures neuves poursuit sa chute
- Pays de la Loire** 08:50 **Birmanie - Politique**  
Birmanie. Aung San Suu Kyi évoque le début d'une « nouvelle ère »
- Sport** 07:51 **France - Multimédia et nouvelles technologies**  
Nouvel iPad : faut-il craquer ?

1 / 3 + Toutes les infos

**jactiv.ouest-france.fr**  
Votre site d'infos mais pas que...

## Eva Joly hospitalisée après une chute

PARIS faits divers



## Birmanie. Aung San Suu Kyi évoque le début d'une « nouvelle ère »

BIRMANIE politique



### La boussole présidentielle

Pour qui voteriez-vous?

Présidentielle 2012  
**-16**  
ouest-france.fr/presidentielle2012

Nos applis

**TOUT OUEST-FRANCE**

- Le journal papier
- Abonnement
- Le journal numérique
- Journal en PDF
- Suppléments gratuits
- Offre pro
- Archives et articles
- L'Ouest-Éclair

**ANNONCES**

- Emploi et formation
- Immobilier
- Automobile
- Autres sites
- Obsèques Dans nos cœurs

**THÉMATIQUES**

- Jactiv Ouest-France
- Ouest-France Entreprises
- meilleurenclasse.com



# Rue89Lyon

# Pure players



Les électeurs de 2012 |  
OL land le projet qui fâche |  
Législative sanglante |  
Affaire Neyret |  
Lyon c'est le Sud |  
Néofascistes en force |

POLITIQUE

SOCIÉTÉ

CULTURES

icoAgenda

Par Laurent Bulet | Rue89Lyon | 29 mars 2012 19:36 5 Commentaires



## UN Oeil

Quelle place pour la culture après mai 2012 ?

2 avr 2012

Psychodrame au Conseil régional : Le Schéma Régional Climat Air Energie n'est pas adopté

30 mar 2012

Manif des agents du Département : la séance publique suspendue

30 mar 2012

L'ancien commissaire détournait des fichiers pour faciliter le divorce de sa compagne

29 mar 2012

RECHERCHER

## L'INFO À 3 VOIX

Journalistes, Experts & vous !

Sciences-po Lyon : peut-on encore critiquer l'état d'Israël en milieu universitaire français ?

Par un collectif d'étudiants  
Le 2 mars dernier, la direction de Sciences-po Lyon déclarait la fermeture administrative de



Search input field



Recherche sur le Web

# Portals

## Actualités

Édition France

Moderne

### À la une

- Mohamed Merah
- Newcastle
- Tegra 3
- Trayvon
- Motta
- Paquebot
- Elgin
- Twilight 4
- Henrique Capriles
- The Voice

Île-de-France

International

France

Économie

Science/High-Tech

Culture

Sport

Santé

Gros plan

### À la une

#### Birmanie : et maintenant, quel rôle pour Aung San Suu Kyi ?

Le Point - Il y a 9 minutes

La "Dame" vient d'obtenir un siège de député, selon son parti. Quel avenir politique attend maintenant Aung San Suu Kyi ? La Birmanie se découvre une scène politique inédite avec l'élection - revendiquée par son parti - de l'opposante Aung San Suu Kyi.

Aung San Suu Kyi salue le «début d'une nouvelle ère en Birmanie» Libération

Aung San Suu Kyi célèbre le triomphe du peuple birman Reuters

Zone géographique : Haïti : Birmanie. Aung San Suu Kyi évoque le début d'une « nouvelle ère » Ouest-France

Afficher toutes les sources (532) »

Autres

Aung San Suu Kyi »

Birmanie »



Europe1

#### Sarkozy confirme 115 milliards d'euros d'économies

Europe1 - Il y a 30 minutes

Nicolas Sarkozy demandera bien 115 milliards d'euros d'économies aux Français. Le président-candidat a confirmé ce chiffre dans une interview publiée lundi dans trois journaux régionaux, l'Est Républicain, le Républicain Lorrain et Vosges Matin.

#### Hospitalisée dimanche, Eva Joly devrait sortir lundi

La Croix - Il y a 22 minutes

La candidate écologiste à la présidentielle a fait une chute dans un escalier alors qu'elle sortait d'un cinéma parisien et a été blessée à la tête. Eva Joly dans les locaux de France Info dimanche 1er avril. Selon son entourage, la participation à ...

#### Analyse du match Rennes-Lyon !

Football.fr - Il y a 15 minutes

Le match du dimanche soir sur Canal était Rennes qui recevait Lyon (1-1). Retour sur le match et les deux équipes. Commençons par l'équipe qui reçoit à savoir Rennes qui a fait un très bon match notamment en première mi-temps.

#### Crash d'un avion en Sibérie : 32 morts, 11 blessés

Libération - Il y a 57 minutes

Un avion ATR-72 s'est écrasé lundi en Sibérie (Russie) avec 43 personnes à bord, faisant 32 morts et 11 blessés. (Photo Frederick Florin. AFP) L'avion, un ATR-72, transportait 39 passagers et 4 membres d'équipage. Il s'est écrasé à 45 km de la ville de ...

#### Sidaction : Les chiffres à retenir de l'édition 2012

France Soir - Il y a 1 heure

France Soir

L'édition 2012 du Sidaction est terminée même si le 110 reste ouvert. 4,2millions d'euros ont été récoltés. Un bon résultat. Voilà les chiffres à retenir. C'est le montant des promesses de dons au profit de la recherche contre le sida, ...

Personnaliser Goog

### Articles récents

#### La Cédéao se réunit, évoque la crise ma

Le Monde - Il y a 12 minutes

#### Aung San Suu Kyi c du peuple birman

Reuters - Il y a 14 minutes

#### Arnaud Montebourg quoi», estime Jean-L

20minutes.fr - Il y a 18 min

Île-de-France » - N

#### Porte de Versailles, N séduit plus la jeunesse

Marianne2.fr - Il y a 56 mir

#### Les riverains divisés

Le Parisien - Il y a 16 minu

#### « Nicolas Dupont-Aigr dire »

Le Parisien - Il y a 7 minut

# Sarkozy proposera 115 milliards d'euros d'économies

Mots clés : Élection Présidentielle, Candidat PS, Position Des Candidats, Candidat UMP, France, Nicolas Sarkozy, François Hollande

Par  Mathilde Golla

Mis à jour le 02/04/2012 à 10:29 | publié le 02/04/2012 à 06:42  Réactions (300)



Le candidat sortant chiffre son programme à 115 milliards d'euros d'efforts demandés aux Français Crédits photo : POOL/REUTERS

 Recommander

**Le candidat à sa propre réélection qui doit détailler son programme en fin de semaine vise 75 milliards d'euros d'efforts sur les dépenses et 40 milliards sur les recettes afin de ramener fin 2016 le déficit à zéro.**

**A 20 jours du premier tour de l'élection présidentielle, le président-candidat dévoile des détails sur le chiffrage de son programme électoral.** A l'occasion de sa venue en Lorraine ce lundi et avant de **publier son programme dans sa totalité en fin de semaine**, Nicolas Sarkozy a ainsi indiqué que l'ensemble de son programme demandera 115 milliards d'euros d'efforts aux Français, dans une interview publiée par trois journaux régionaux l' *Est Républicain* , le *Républicain Lorrain* et *Vosges Matin* .

# Getting context and comments



michel brillaud

Quel est le plan d'économies de Hollande ? Mystère...  
Qu'il nous le présente maintenant puisque tout le monde sait que l'on y coupera pas, au lieu de finasser et de gagner du temps .

Le 2/04/2012 à 10:17

 Répondre  Alerter



parachute

DE qui se moque-ton! Si on l'écoute pour les recettes tout est déjà voté. Quand au massacre sur les services publics pour 75 milliards si c'est de la même veine que les économies de SS que l'on paye en augmentation des mutuelles....

Le 2/04/2012 à 10:07

 Répondre  Alerter



grappaloup

10 % (dix pour cent) de la population active paie 70 % (soixante dix) de l'impôt sur le revenu.

Les 90 % (quatre vingt-dix) ne paient que les 30 % (trente) restant !  
Qui peut prétendre que les riches ne paient pas beaucoup d'impôts ?  
La municipalité d'Aubagne est la ville de France la plus endettée à hauteur de plus de 125 millions d'euros. Elle a du faire appel aux capitaux de la finance et accepter qu'une agence de notation vienne l'auditer (vérifier ses comptes et budgets) !

Ah ! j'oubliais : cette ville est dirigée par des communistes, style Mélanchon qui crache sans arrêt sur les riches, la finance, les agences de notation !

Le 2/04/2012 à 10:05

 Répondre  Alerter



YAN A MARRE

Les français ont compris depuis longtemps qu'il faudrait se serrer la ceinture. La question est de savoir qui sera le plus mis à contribution. Les 5 ans qui viennent de s'écouler sont déjà un élément de

{BnF



# Insight into BnF collections

Some issues

ÉDITION: FR

PlusConnexion | Créer un compte

- [élection présidentielle russe](#)
- [Présidentielle 2012](#)
- [Tendances](#)
- [états-unis](#)

# HUFF POST PRÉSIDENTIELLE 2012

J'aime 15k

Rechercher sur le HuffP

PRÉSIDENTIELLE 2012

EN CE MOMENT [CRISE DE LA DETTE](#) [NICOLAS SARKOZY](#) [FRANÇOIS HOLLANDE](#) [RÉPRESSION EN SYRIE](#) [IMPÔTS](#)  
[APPLE](#) [ÉLECTIONS AMÉRICAINES](#) [LA VIE DIGITALE](#) [PEOPLE](#)

Recevoir des alertes

Flux RSS de cette page

J'aime 22

# Loss of original form

## Nathalie Arthaud a ses 500 signatures

# OU EN SONT LES AUTRES



{BnF



The  
website  
online

ÉDITION: FR **Sport** Présidentielle 2012 Économie La Vie Digitale Plus Connexion | Créer un compte

HUFF POST PRÉSIDENTIELLE 2012

J'aime 17 k Suivre +1 Rechercher sur le HuffPost CONNECTEZ-VOUS

À LA UNE PRÉSIDENTIELLE 2012 ÉCONOMIE INTERNATIONAL CULTURE TENDANCES

EN CE MOMENT CRISE DE LA DETTE NICOLAS SARKOZY FRANÇOIS HOLLANDE RÉPRESSION EN SYRIE IMPÔTS APPLE ÉLECTIONS AMÉRICAINES LA VIE DIGITALE PEOPLE

Entrez votre adresse email Recevoir des alertes

69€ frais de port inclus Les plus grands films de l'histoire du cinéma Je commande le coffret

LE CINEMA DU Monde

J'aime 25 FLUX RSS DE CETTE PAGE

# DIX CANDIDATS



MEDIAPART

Rechercher  ok

**9€**  
**ABONNEZ VOUS ICI** ▶

CONNECTEZ-VOUS

Identifiant

OK

Mot de passe oublié ?

LE JOURNAL

LE CLUB

MODE D'EMPLOI

**LA LUNE**

**Dans le quartier du Prepaou, à Istres, «la présidentielle, c'est du marketing»**

- ▶ Dans l'Essonne, Mélenchon veut...
- ▶ Pour les banlieues, Hollande veut créer un grand ministère de l'égalité
- ▶ 2007-2012: le mirage de la «France de la croissance»
- ▶ L'air de Sarkozy
- ▶ Dupont-Aignan: «Le système veut m'assimiler au FN»
- ▶ L'œil de Patrick Artinian: Nicolas Sarkozy, les jeunes et le «droit des victimes»
- ▶ Pinocchio, conte défait
- ▶ Le courant alternatif de la poésie peut-il se faire politique?
- ▶ La guerre économique version Casino
- ▶ Le couple syro-iranien

**Testez-nous pour 1€ seulement !**  
 Profitez d'une période d'essai de 15 jours  
 + 15 jours de musique offerts ! [Cliquez ici](#)

ÉDITION : JAPON, UN SÉISME MONDIAL

**Elle est pas belle la vie ? :  
 "Effacer Fukushima"**

PAR CORINNE N

Je vous propose la lecture du dernier billet de Pierre Fetet, intitulé "Effacer Fukushima", dans lequel il fait la démonstration des mensonges et du déni de réalité des instances gouvernementales accompagnées par les médias au sujet de la catastrophe de Fukushima.

[Lire la suite](#)

**Tout fout le camp !**

PAR JOËLMARTIN

Les certitudes les plus absolues s'effondrent. Etoile66 est devenue germanophobe. JNSPQD, Dianne, Penelope, Simeoni, Padovani et Tutti Quanti ont viré ségolénophobes.

[Lire la suite](#)

ÉDITION : LES INVITÉS DE MEDIAPART

**L'aide publique au développement dévoyée**

PAR LES INVITÉS DE MEDIAPART

Philippe Marchesin, spécialiste des relations Nord-Sud et de l'aide au développement à l'Université Paris 1, décrypte le budget de l'«APB» tel que soumis au Parlement: «opaque», «insincère» et où les besoins des pays censés en être destinataires sont sacrifiés «sur l'autel du commerce et de la sécurité».

[Lire la suite](#)

LES ÉDITIONS RECOMMANDÉES

PAR LA RÉDACTION PAR LES ABONNÉS



**Je me souviens...**  
 Dernière vidéo : Si j'étais président...  
 PAR CHRISTOPHE LASTERLE



**Contes de la folie ordinaire**  
 Dernier article : La folie, la chercheuse de Varsovie et les «scient ...»  
 PAR HEITOR O'DWYER DE MACEDO



**La mendie-bulles**  
 Dernier article : Un collage...  
 PAR ALEXIS FLANAGAN



**Complices**  
 Dernier article : L'économie doit sortir de la pensée unique  
 PAR LA RÉDACTION DE MEDIAPART



LES BLOGS RECOMMANDÉS

PAR LA RÉDACTION PAR LES ABONNÉS



**chroniques**  
 PAR MARIE COSNAY

Protected content

Je m'identifie



Je m'abonne à partir de 1€ seulement

Newsletter



Lun.  
02  
Avr

# MEDIAPART

Accédez aux archives



CONNECTEZ-VOUS




Mot de passe oublié ?

LE JOURNAL INTERNATIONAL FRANCE ECONOMIE CULTURE ENGLISH

LE CLUB

LA UNE

Dans le quartier du Prepaou, à Istres, «la présidentielle, c'est du marketing»

## Dans le quartier du Prepaou, à Istres, «la présidentielle, c'est du marketing»

02 AVRIL 2012 | PAR RACHIDA EL AZZOUI

« Ça sert à rien de voter. Ils viennent nous voir uniquement lorsqu'ils ont besoin de nous au second tour, et après il n'y a plus personne. » C'est le refrain qui tourne en boucle, d'immeuble en

immeuble, au Prepaou, à moins de vingt jours du premier tour.

Dans ce quartier populaire de la ville d'Istres, dans les Bouches-du-Rhône, la masse de précaires a perdu la foi en la République et en ses hommes politiques.

**Inscrivez-vous pour avoir accès à cet article immédiatement.**



**1€ SEULEMENT  
LES 15 PREMIERS JOURS**

**TESTEZ-NOUS !**

**VOUS ÊTES ABONNÉ(E)  
IDENTIFIEZ-VOUS**

Identifiant ou e-mail : \*

**JE M'INSCRIS**

Choisissez votre formule et créez votre compte pour accéder à tout Mediapart

- ▶ Dans l'Essonne, Mélenchon veut mobiliser les quartiers populaires
- ▶ Pour les banlieues, Hollande veut créer un grand ministère de l'égalité
- ▶ 2007-2012: le mirage de la «France de propriétaires»
- ▶ L'affaire Sarkozy
- ▶ Dupont-Aignan: «Le système veut m'assimiler au FN»
- ▶ L'œil de Patrick Artinian. Nicolas Sarkozy, les jeunes et le «droit des victimes»
- ▶ Pinocchio, conte défait
- ▶ Le courant alternatif de la poésie peut-il se faire politique?
- ▶ La guerre économique version Casino
- ▶ Le couple syro-iranien

### MOTS-CLÉS

banlieues quartiers populaires  
présidentielle 2012  
Bouches-du-Rhône Istres  
Prepaou

### DERNIERS ARTICLES

- ▶ Dans l'Essonne, Mélenchon veut mobiliser les quartiers populaires
- ▶ L'œil de Patrick Artinian. Nicolas Sarkozy, les jeunes et le « droit des victimes »
- ▶ L'affaire Sarkozy
- ▶ Six ans requis contre Adlène Hicheur, « terroriste en puissance »
- ▶ Le beau rôle libyen de Cécilia Attias

### DU MÊME AUTEUR

- ▶ Jeunes profs, vocations brisées
- ▶ Kadhafi-Sarkozy: Eva Joly demande l'ouverture d'une enquête judiciaire
- ▶ Le storytelling de Sarkozy écrase les affaires
- ▶ Formation: l'AFPA plongée dans une crise «à la France Télécom»
- ▶ Le PS se divise sur le «MES» à l'Assemblée





## Going further : the “press project”

- Currently performed with Ouest France
  - The 50 local editions of Ouest France are not gathered anymore in paper form
- On the harvesting side
  - Giving the password to the robot in order to let it capture protected content
  - Collecting PDF equivalents of printed versions
- On the access side
  - Making a link from the catalogue record of the print version, to the archived PDF version
- Results expected in few months (more information in Mikkeli !)



## To conclude

- Collecting online newspaper websites is key to ensuring the continuity of heritage collections
- For BnF, legal framework is provided by legal deposit
- Web crawlers represent an inexpensive way to gather a large number of collections
  - But some technical issues remain
- More complex harvesting or deposit operations may be necessary in order to gather protected content

{BnF



Thank you for your  
attention!

