



# Actual and potential users of the BnF web archives



*Experiences and expectations*

Clément Oury and Peter Stirling

Department of Digital Legal Deposit  
Bibliothèque nationale de France

# Summary



- Experiences:
  - The context: access to the BnF web archives
  - Who consults the web archives?
  - What are they looking for?
- Expectations:
  - How do non-users imagine the web archives?
  - What services do they look for?
  - How can we respond to users' needs?

# Access to the BnF 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 since 2009

# Access to the BnF web archives

- Three ways of consulting archives:
  - URL searching
  - (Very limited) full-text searching
  - « Guided tours », selections of sites by librarians

## Use cases of the web archives

- Consultation remains stable but limited: 30-50 visits per month
- Use cases noted during the accreditation process in 2008-2010
- Mixture of research projects, individual questions and simple curiosity

# What are users looking for?

- Researchers in political and social sciences are the most frequent visitors
- Many researchers studying election campaigns and the role of the Internet
  - In particular Presidential elections in 2002 and 2007 : blogs, caricatures...
  - European referendum in 2005 and European elections in 2009
  - Much of this material no longer existed online

# What are users looking for?

- Other research themes:
  - Representation of housing projects in the media, in video games and online
  - Reactions to riots in 2008
  - Museum access for disabled people
  - Amateur writers' websites
  - Marketing, for example craftspeople selling their work online

# What are users looking for?

- Professional or personal searches:
  - How to start an Internet project
  - Web design and graphic design
  - Health information on the Internet
  - Website of a ceramics factory...



## Users' reactions

- Positive:
  - Most users found what they were looking for and were pleased to discover this resource
  - Several users said they found the archives clear and simple to use
- Negative:
  - Impossible to copy files for further use
  - Missing content, or difficulty in accessing content (Flash, videos...)
  - Lack of information on the reason for collecting a given URL on a given date
  - Lack of information on quality of a given capture

# Potential users of BnF web archives

- Study in late 2010/early 2011 based on in-depth interviews with 15 individuals who had never used the web archives
- Five researchers, five professionals and five average users of the library
- Aims:
  - Examine practice in research on the web
  - Explore perceptions of web archiving
  - Identify new approaches to constituting archives and new tool and services

## ● ● ● | Researchers and the web

- Researchers use the web but often as a way of exploring new areas or communicating with peers
- Citing web-based material still seen with suspicion:
  - How to define what is important?
  - How to cite stable, « authentic » sources?

# Perceptions of web archiving

- Value of maintaining the « memory » of the web clear to researchers
- Ethical questions: public/private
- Problems with imagining or representing a web archive:
  - The web is fluid, seems opposed to idea of a « static » archive
  - Some confusion with « archives » on websites
  - Paper archive is not necessarily a model: web archiving more like the oral tradition



# Content and selection

- Selection of some kind necessary faced with the mass of data on the web
- To make a meaningful archive, the choices made must be clear and fully described
- Various approaches proposed
  - by theme, by what is « groundbreaking », by popularity, by Google search results...

# Information needs and potential services

- Document the collection process and describe the archive
  - Vital for understanding the nature of the archive
- Cooperation with communities
  - Work with researchers to improve focused crawling
- Provide tools for data analysis
  - Beyond searching researchers need to analyse data
  - Researchers may already have tools that could be used in the BnF (in the limits of what the law allows)

# What we do well... and less well

- A lot of content (from 1996-present, 50-80 Tb each year since 2010)
- Well-defined selection policy
  - ...but not visible in the archives
- Access to all research users
  - ...but only on-site
  - ...no possibility of reusing files
  - ...no permanent links for citation
- Search by URL for specific site
  - ...but no full-text searching
  - ...no tools for link analysis, semantic analysis

# Changes in access to the BnF web archives

- With publication of decree access is no longer “experimental”
- Programme of improvements in the access interface begins this year
- Access will be opened in main regional libraries





# References

- Stirling, Chevallier, Illien “Web Archives for Researchers: Representations, Expectations and Potential Uses”. In *D-Lib Magazine*, vol. 18(3/4), March/April 2012  
<http://www.dlib.org/dlib/march12/stirling/03stirling.html>
  - Article concentrating on interviews with five researchers
- Chevallier and Illien. *Les archives de l'Internet: une etude prospective sur les representations et les attentes des utilisateurs potentiels*. Bibliothèque nationale de France, 2011  
[http://www.bnf.fr/documents/enquete\\_archives\\_web.pdf](http://www.bnf.fr/documents/enquete_archives_web.pdf)
  - Report on the full study (in French)

{BnF



Questions ?



{BnF



Screenshots of BnF  
web archives interface



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>).




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).




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.



Avertissement : les techniques utilisées ne...

Retrouver ce site ou cette page :

**Remonter le temps**

**Option**

Limiter à :

- - -

## Recherche par URL

pour : http://www.lemonde.fr

1 084 résultats

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 ▶

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

pour : http://www.lemonde.fr

1 084 résultats

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



Recherchez depuis 1 mois

Identifiez-vous email Mot de passe

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

# 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 rannel. niveau faible

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

- 18h11 **Portfolio** "Il n'y a pas de suspense"
- 16h15 **Vous** Vos photos du second tour de la présidentielle
- 17h05 **Portfolio** A Belleville, les électeurs de François Bayrou à l'heure du choix
- 15h17 **Vidéo** Vote électronique à Issy-les-Moulineaux
- 16h39 **Infographie** Taux de participation à la mi-journée pour le second tour de l'élection présidentielle depuis 1969
- 15h06 | Réactions (8)
- Les faits** Une forte mobilisation policière prévue pour le second tour de la présidentielle



Recherche 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?

# 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 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

## 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





Recherchez depuis 1 mois sur Le Monde.fr sur le web avec YAHOO! SEARCH

Identifiez-vous email Mot de passe \*\*\*

# 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 LES RÉSULTATS 2d tour Votre commune ou votre code postal Votre département

### 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

### 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



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 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.



### (S') écrire en ligne : journaux personnels et littéraires

Ce dossier montre comment le passage du papier à l'écran et la forme du blog ont transformé le journal intime, mais aussi l'écriture et la critique littéraires.



**Avertissement :** les techniques utilisées ne



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

Retrouver les occurrences de ces mots :

**Rechercher****Possibles :**

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

**” Recherche par mot**

495493 résultats

pour : ministère développement durable

résultats 1 à 10 -  par page

1 2 3 4 5 6 7 8 9 10 ▶▶

**Comité français pour l'environnement et le développement durable - Comité 21**<http://www.comite21.org/index.php>

... Magazine : Des réseaux haute fidélité - juin 2006 20 minutes : Agenda 21, la France enfin sur son trente et un Le **Ministère** de l'écologie et du **développement durable** vient de mettre en ligne la Stratégie Nationale de **Développement Durable** 2003-2008 actualisée composée de trois documents : 1 - Objectifs stratégiques et instruments 2 - Programme d'actions 3 - 12 indicateurs "phares ... pour la création d'une organisation mondiale de l'environnement (ONU) organisée par le Comité 21, à la demande du **Ministère** de l'écologie et du ...

Archive du 16/11/2006 à 19:03 | autres dates

html - 53 Ko

**MEDAD - Sites Transports et Mer**<http://www.transports.equipement.gouv.fr/>

... et Mer Aller au contenu Retour à l'accueil Présentation Presse Voyageurs Marchandises Social et Économie Sujets transversaux Nos priorités **Développement durable** Sûreté et sécurité des transports Transports : dimension sociale et économique Grands projets L'intermodalité Liens utiles Inspection du travail des ... International Economie et Statistiques Réseau ferré de France SNCF RATP Voies navigables de France Légifrance La documentation française Vie publique **Ministère** de l'Écologie, du **Développement**

Archive du 12/10/2007 à 05:06 | autres dates

html - 32 Ko

**Ministère du développement durable**<http://www.bougezautrement.gouv.fr/>**Ministère du développement durable**

Archive du 12/10/2007 à 00:47 | autres dates

html - 0 Ko

**Les publications du ministère de l'écologie et du développement durable : publications**



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

[Tous les parcours](#)[Le web militant](#)[\(S\) écrire en ligne :  
journaux personnels et  
littéraires](#)[Cliquer, voter : l'Internet  
électoral](#)

## Parcours guidé

Créé le 10/03/2008

### Cliquer, voter : l'Internet électoral

Les sites des acteurs, observateurs et témoins des campagnes électorales. Ce dossier couvre et compare les élections présidentielles et législatives de 2002 et 2007 ainsi que les élections régionales et européennes de 2004.



### À propos de ce parcours...

Ce parcours permet de découvrir les archives constituées lors des campagnes électorales de 2002 à 2007. Les élections concernées sont :

- en 2002 : l'élection présidentielle (21 avril et 6 mai) et les élections législatives (9 et 16 juin),
- en 2004 : les élections régionales (21 et 28 mars) et les élections européennes (13 juin),
- en 2007 : l'élection présidentielle (22 avril et 6 mai) et les élections législatives (10 et 17 juin).

Les sites collectés ont été repérés par des bibliothécaires de la BnF. Pour les élections locales (régionales et européennes de 2004, législatives de 2007), le travail de repérage a été élargi à d'autres établissements : en 2004, les Bibliothèques de dépôt légal imprimeur de Lyon, Caen, Rennes et Toulouse. En 2007, Dijon, Lille, Limoges, Nouméa, Poitiers et Strasbourg sont venus rejoindre Caen et Lyon.

La part de plus en plus importante prise par l'Internet politique dans le processus électoral a

### Thèmes de ce parcours

- Être candidat à une élection politique
- Les "petits candidats" aux élections présidentielles de 2002 et de 2007
- Les candidatures aux candidatures PS et UMP pour l'élection présidentielle de 2007
- Élection présidentielle 2007 : à la recherche d'une candidature unitaire de la gauche anti-libérale
- La campagne de Nicolas Sarkozy à l'élection présidentielle de 2007
- La campagne de Ségolène Royal à l'élection présidentielle de 2007
- La campagne de François Bayrou aux élections présidentielle et législatives de 2007
- La campagne de Jean-Marie Le Pen à l'élection présidentielle de 2007
- La campagne en région : l'exemple de la région Rhône-Alpes
- La France politique d'Outre-mer à portée de clic
- L'écologie : de la niche politique au débat de

Tous les parcours

Le web militant

(S) écrire en ligne :  
journaux personnels et  
littéraires

Cliquer, voter : l'Internet  
électoral



## Parcours guidé

### Cliquer, voter : l'Internet électoral

#### Les boîtes à outils pour les militants

Les militants sont appelés par leurs partis à s'impliquer dans le débat électoral. De nombreuses boîtes à outils sont mises à leur disposition pour argumenter, manifester leur soutien et même créer leur propre site d'opinion.

#### La rubrique "Militez" du site de campagne du Mouvement pour la France en 2002

<http://www.mpf-villiers.org/Militez/index.html>

► Archive du 04/04/2002 à 09:00 | autres dates



#### Les "outils de campagne" du site de Christiane Taubira en 2002

<http://www.christianetaubira.org/asp/static7887.html?page=ht>

► Archive du 16/04/2002 | autres dates



#### Le site de l'UMP en 2004 propose d'acheter des produits estampillés UMP pour manifester son soutien

[http://www.u-m-p.org/e\\_boutique/index.html#](http://www.u-m-p.org/e_boutique/index.html#)

► Archive du 25/02/2004 | autres dates



#### Argumentaires et affiches sur le site de campagne à l'investiture socialiste de Dominique Strauss-Kahn



## Thèmes de ce parcours

- Être candidat à une élection politique
- Les "petits candidats" aux élections présidentielles de 2002 et de 2007
- Les candidatures aux candidatures PS et UMP pour l'élection présidentielle de 2007
- Élection présidentielle 2007 : à la recherche d'une candidature unitaire de la gauche anti-libérale
- La campagne de Nicolas Sarkozy à l'élection présidentielle de 2007
- La campagne de Ségolène Royal à l'élection présidentielle de 2007
- La campagne de François Bayrou aux élections présidentielle et législatives de 2007
- La campagne de Jean-Marie Le Pen à l'élection présidentielle de 2007
- La campagne en région : l'exemple de la région Rhône-Alpes
- La France politique d'Outre-mer à portée de clic
- L'écologie : de la niche politique au débat de société
- **Les boîtes à outils pour les militants**
- La publication des résultats d'une élection
- 21 avril 2002 : réactions aux résultats du premier tour de l'élection