

## Dossier de presse

# Internet en campagne

## de l'archivage des sites électoraux à leur communication au public

---

*Après le vote de la loi instaurant le dépôt légal de l'Internet et en attendant la parution du décret d'application, la Bibliothèque nationale de France entreprend de collecter et d'archiver des sites électoraux, d'octobre 2006 à juillet 2007, dans la continuité des expérimentations déjà conduites en 2002 et 2004 à l'occasion des élections précédentes.*

*Le corpus recueilli donnera lieu à un travail sur l'usage futur de ces précieuses sources nouvelles en partenariat avec Sciences Po.*

<b>Communiqué de presse</b>	<b>2</b>
<b>L'archivage de la campagne électorale sur Internet</b>	<b>3</b>
<ul style="list-style-type: none"><li>• En 2007, pour la première fois, l'Internet sera véritablement en campagne</li><li>• Les principes documentaires retenus par la BnF pour la campagne d'archivage des sites électoraux</li></ul>	
<b>L'étude d'usages en partenariat avec Sciences Po</b>	<b>6</b>
<ul style="list-style-type: none"><li>• Nouveau média, nouvelles sources : les problématiques abordées</li><li>• Principes et organisation pratique de l'étude</li><li>• Le corpus de l'étude : archives des campagnes électorales de 2002 et 2004</li></ul>	
<b>Glossaire</b>	<b>10</b>

---

## **Communiqué de presse**

**Dans la perspective des élections de 2007, la Bibliothèque nationale de France, lance deux projets liés à la vie politique nationale qui s'inscrivent dans le sillage de la mise en place du dépôt légal de la Toile.**

**C'est l'occasion d'illustrer les innovations technologiques et documentaires à l'œuvre au sein de l'établissement, dans la continuité de ses missions et métiers.**

Le premier projet porte sur la collecte de sources nouvelles pour l'histoire politique et sociale de la nation ainsi que pour la sociologie électorale. Pendant dix mois (d'octobre 2006 à juillet 2007), la BnF archivera des sites électoraux nés à l'occasion de l'élection présidentielle et des élections législatives de l'année 2007.

Ce projet s'inscrit dans la continuité des expérimentations déjà conduites par l'établissement lors des processus électoraux de 2002 et 2004 (la BnF avait alors collecté environ 3600 sites au total). Il s'agira cette fois d'un ultime test de collecte événementielle avant que la Bibliothèque ne s'engage dans des captures de plus grande échelle, conduites à des fréquences plus rapprochées. Alors que l'on observe un foisonnement sans précédent des sites relatifs à la vie politique, l'opération permettra ainsi de tester l'organisation documentaire et les performances techniques du dispositif de moissonnage automatique dans la perspective d'une montée en charge progressive de cette activité, en application de la loi du 1<sup>er</sup> août 2006 qui confie à la BnF cette mission nouvelle de dépôt légal de l'Internet.

Le deuxième projet concerne la consultation en Rez-de-Jardin par les chercheurs accrédités des archives d'Internet. En attendant la parution du décret qui l'autorisera à communiquer les archives collectées par ses soins, la BnF s'apprête à lancer un chantier qui n'est pas seulement technique : organiser les accès aux sites archivés depuis ses salles de lecture, imaginer et tester les outils, les services et les conditions de consultation capables de répondre aux futures attentes des chercheurs qui se verront bientôt offrir des formes inédites de documentation.

Une étude d'usages a donc été entreprise en partenariat avec Sciences Po, pôle associé de la BnF, autour des archives des campagnes électorales de 2002 et 2004. La Bibliothèque suivra de près les travaux de recherche et les parcours d'apprentissage d'un groupe d'étudiants et de chercheurs en sciences politiques qui seront tout à la fois les cobayes et les observateurs critiques du dispositif.

### **Contact presse :**

Claudine Hermabessière, chargée des relations avec la presse

Tel : 01 53 79 41 18 - [claudine.hermabessiere@bnf.fr](mailto:claudine.hermabessiere@bnf.fr)

## L'archivage de la campagne électorale sur Internet

En 2007, pour la première fois l'Internet sera véritablement en campagne

La campagne électorale de 2007 a déjà commencé et Internet pourrait y jouer un rôle jamais égalé à ce jour. Les acteurs de la campagne, formations constituées ou communautés virtuelles issues de la société civile, structurent l'espace politique d'Internet en même temps qu'Internet les structure et influe sur leurs stratégies.

Une visite des principaux sites des **organisations politiques et des candidats** révèle une nette évolution de leur rapport à Internet : utilisation plus répandue du multimédia, de la baladodiffusion, mise à disposition de kits de campagne téléchargeables, de forums, de *chats*, d'espaces et de techniques d'édition invitant militants et sympathisants à l'écriture collaborative, possibilité d'adhérer en ligne, ou encore techniques plus agressives de publipostage ou d'achat de mots clés. Les blogues constituent également un vivier de libre parole sans précédent qui, depuis les élections de 2004, n'a cessé de croître pour devenir un mode d'expression à part entière.

Comme pour les élections précédentes, **les observatoires et sites d'information sur les élections sont également nombreux**. Très visibles, ces portails sont le plus souvent en tête des interrogations par mots clés significatifs. Ils peuvent être considérés comme des intermédiaires chargés de mettre en forme une offre bientôt pléthorique de sites politiques. Ils proposent des illustrations cartographiques et des interprétations statistiques basées, par exemple, sur le référencement et la notoriété des sites. Vu depuis ces nouveaux instituts de sondage virtuels, le site devient un objet d'étude à part entière, entièrement légitimé par sa forme.

Garder la trace, « en temps réel », de ces innovations radicales constitue l'un des enjeux majeurs de la campagne d'archivage conduite par la BnF.

## Les principes documentaires retenus par la BnF pour la campagne d'archivage des sites électoraux

Le premier axe de sélection se concentre sur **les candidats et leurs organisations**. Ce premier ensemble s'inscrit dans la continuité de la tradition documentaire de la BnF en science politique et en histoire (recueils, tracts...). La Bibliothèque collectera de façon systématique les sites des partis politiques, mais aussi les sites et blogues officiels ou personnels des candidats en campagne.

Le deuxième axe, intitulé **regards et opinions sur la campagne**, rassemble un ensemble d'illustrations et de traces des réactions et contributions au débat des citoyens et, plus largement, de la société civile. Il devrait permettre de retrouver sur Internet des acteurs collectifs « classiques » des campagnes (comités de soutien, syndicats, autres organisations constituées intervenant à des titres divers dans le débat public), mais aussi d'explorer des formes émergentes d'expressions individuelles et de réseaux d'opinion qui n'existent que sur la Toile. Compte tenu de la profusion et de la volatilité des sites dans ce domaine, la BnF ne pourra proposer qu'un échantillon d'exemples, qui couvrira un spectre aussi large que possible, sans prétendre à une représentativité parfaite.

La BnF s'attachera à traduire toutes les sensibilités politiques dans un souci de neutralité et de représentativité. **Elle entend aussi représenter le double niveau national et local**. Le scrutin législatif (juin 2007), qui fera suite au scrutin présidentiel (mai 2007), sera l'occasion de porter une attention particulière à des sites de campagne et d'expression politique reflétant les enjeux dans les différentes régions de France. Le principe d'une reconduction de la collaboration menée en 2004 avec les bibliothèques de dépôt légal imprimeur, pôles associés, et d'autres bibliothèques territoriales a été retenu.

## *Typologie documentaire retenue pour la sélection des sites*

- **Sites officiels et institutionnels** : règles du jeu, cadre juridique et réglementaire
- **Les candidats et leurs organisations** :
  - **Sites des candidats en campagne** : sites et blogues officiels ou personnels, permanents ou de campagne, des candidats
  - **Sites des formations politiques** : partis et formations en campagne
  - **Autres organisations de soutien** : uniquement les sites clairement affiliés aux partis et aux candidats : comités de soutien par exemple
- **Regards et opinions sur la campagne** :
  - **Annuaire, observatoires et analyses** : y compris les sites de sondages et de tendances, les nouveaux outils d'analyse de type cartographique, « bruitomètres »...
  - **Médias traditionnels** : sélection restreinte à un nombre limité de titres de presse d'information disposant d'un site à l'activité significative
  - **Associations, syndicats et autres organisations** : cette catégorie doit servir à relayer la parole d'acteurs collectifs qui n'existent pas uniquement sur Internet
  - **Expressions individuelles et communautaires sur Internet** : cette catégorie vise à regrouper les formes d'expression et d'agrégation de l'opinion à la fois structurées par le média Internet et structurantes de l'organisation et des usages de la Toile

L'infrastructure utilisée pour réaliser cette collecte s'appuie sur le robot Heritrix<sup>1</sup>, avec six machines fonctionnant simultanément 24h sur 24.

<sup>1</sup> Documentation logicielle accessible en ligne depuis le site : <http://crawler.archive.org>

## **L'étude d'usages en partenariat avec Sciences Po**

La BnF a proposé à la Fondation Nationale des Sciences Politiques (bibliothèque de Sciences Po – pôle associé Sciences Po) d'être partenaire du projet d'étude d'usages conduit d'octobre 2006 à décembre 2007.

Cette étude a pour objectifs de **tester les interfaces** existantes et d'observer des **situations d'apprentissage et de recherche**. Elle vise ainsi à valider la pertinence des modalités d'accès imaginées et à préfigurer les conditions de consultation futures de ces archives.

Concrètement, ce « galop d'essai » peut aider la Bibliothèque à concevoir **des outils et des services utiles au chercheur** de demain. En outre, le partenariat avec Sciences Po vise à **élargir la réflexion à des problématiques de fond pour les sciences humaines en général**, autour de quelques points essentiels.

### **Nouveau média, nouvelles sources : les problématiques abordées**

Les archives d'Internet, même réduites à leur portion « nationale » (les sites du .fr et d'autres domaines intéressant la France), constituent un défi du fait de leur masse, de leur structure hypertextuelle, de leur temporalité singulière. Elles offrent des possibilités nouvelles d'exploration et d'analyse : l'étude d'un réseau ou d'un tissu d'objets dans leur contexte d'ensemble, et non plus de manière isolée, présente un intérêt scientifique et patrimonial majeur. L'accès simultané à des strates d'existence successives d'un site à partir d'une même interface constituera un mode d'investigation inédit pour l'historien. Enfin, les archives d'Internet donneront accès à des types de documents jamais réunis jusque là et qui illustrent de nouveaux modes d'interventions et d'échanges dans l'espace public tels que les blogues et les espaces d'expression collaboratifs.

### ***Quels outils de mise à disposition de ce nouvel ensemble documentaire ?***

A l'image de la réflexion engagée pour l'ensemble des collections issues du monde numérique, des accès intelligents (« *smart access* », « *data mining* ») faisant appel à la recherche linguistique, à la notoriété des objets ou encore à des modes de représentations visuelles et cartographiques, sont à inventer pour les archives de la Toile et, plus largement, pour la bibliothèque de demain.

### ***Quelle méthode pour les chercheurs face à ce nouvel ensemble ?***

L'archive d'un site réunit un ensemble de captures réalisées à des moments différents : c'est un document noyé dans la masse, lacunaire à bien des titres ; c'est aussi un document relié à d'autres (c'est un « tissu » de sites liés les uns aux autres qu'on archive, et non des unités juxtaposées), ce qui permet de les étudier systématiquement dans un contexte plus large. Comment le chercheur peut-il et doit-il aborder ce nouvel objet du point de vue de la méthode et du point de vue critique ? **Quels sont les changements induits par l'archivage d'Internet ?** Comment comparer ces archives à « l'espace-temps » bien particulier aux autres sources habituellement utilisées en sciences humaines et quelles complémentarités peut-on proposer dans l'exploitation de ces sources ? L'archive électorale offre une étude de cas exemplaire pour engager cette réflexion et s'interroger plus particulièrement sur l'articulation de l'appareil « Internet » avec les autres objets et médias de campagne.

### ***Nouveaux objets d'étude pour les sciences sociales***

**Quel est l'impact d'une campagne « virtuelle » sur les choix des électeurs ?** Est-on capable de la mesurer ? Comment s'y prendre ? Ces questions se posent avec d'autant plus d'acuité qu'avec les pratiques d'expression sur Internet, on ne peut plus se contenter d'un modèle de communication strictement « vertical » ni « horizontal » : ces dynamiques se trouvent intimement mêlées dans un espace par nature participatif et interactif. Les internautes interviennent activement dans la campagne, leurs commentaires viennent s'agréger aux messages des sites officiels ou officieux, les sites personnels et les blogues émanent aussi bien des particuliers que des candidats : les cartes de l'expression et de l'espace publics s'en trouvent radicalement rebattues, et les contenus sont eux-mêmes porteurs d'usages.

### ***La spécificité du média Internet au service des sciences humaines et sociales***

Sur Internet encore plus que sur d'autres supports, ce qui donne sens et poids à l'information n'est plus seulement son contenu mais aussi sa notoriété, son référencement dans et par d'autres sites. L'Internet dit de seconde génération, qui conduit à un développement de ces fonctions de « pointage » (*linkage*) renforce encore ce phénomène. Les modalités de l'archivage de la Toile nous permettent aujourd'hui de mesurer et de représenter ces relations de « pointage » - ou de référencement croisé - qu'entretiennent les sites entre eux. Des outils spécifiques, proches de la cartographie dynamique, sont développés pour proposer des fonctions d'observatoire, de panoramas analytiques permettant de mesurer la position d'un site par rapport à d'autres et de reconstituer par ce biais **des circuits d'influence, des réseaux d'appartenance, d'affiliation, d'opposition ou, plus généralement, de relation entre les acteurs**. La campagne offre un terrain d'exploration particulièrement pertinent pour conduire ces analyses et développer des outils qui tiennent compte de la spécificité forte du média Internet.

## Le corpus de l'étude : archives des campagnes électorales 2002 et 2004

Si l'essentiel des collectes de sites Internet de la BnF est réalisé par des robots lors de campagnes d'archivage massives et automatiques, des collectes thématiques complètent ce dispositif. Elles visent à pallier les insuffisances des robots en s'appuyant sur une prospection documentaire plus ciblée. Il s'agit de signaler manuellement au robot l'adresse des sites à conserver et de lui indiquer selon quelle fréquence et à quelle « profondeur » il doit les capturer. **Des campagnes de collecte peuvent ainsi être organisées autour d'un événement d'intérêt national.** D'autres pays ont, par exemple, choisi de couvrir des événements comme les attentats du 11 septembre 2001 ou, plus récemment, le *tsunami* en Asie. Ce type de collecte est adapté au dépistage de sites événementiels qui surgissent et disparaissent très rapidement.

La BnF a déjà réalisé deux campagnes d'archivage thématique des sites électoraux en France. Environ 1900 sites ont été archivés lors des élections présidentielles et législatives de 2002, et près de 1700 lors des élections régionales et européennes de 2004. Pour les élections régionales de 2004, la participation de plusieurs bibliothèques associées en région a permis de mieux couvrir les intérêts régionaux et locaux.

Le corpus constitué à l'issue de ces deux campagnes est proposé comme objet d'étude dans le cadre de l'étude d'usages. Cette collection présente une unité thématique et chronologique et son mode de constitution et de description facilite sa manipulation, constituant de fait un ensemble au périmètre clairement délimité.

## Principes et organisation pratique de l'étude

Les pratiques interdisciplinaires qui caractérisent la recherche en science politique, l'engagement reconnu de Sciences Po dans des axes de recherche et d'enseignement qui convergent avec ce projet ainsi que l'existence du corpus électoral sont rapidement apparus comme autant d'arguments en faveur de cette association. La BnF a ensuite sollicité et obtenu de la plupart des partis et mouvements politiques éditeurs de sites Internet l'autorisation d'accéder aux archives de leurs sites dans le cadre de cette expérience.

L'étude se compose de trois volets :

- **un volet pédagogique** : organisation et animation conjointe par le pôle associé et la BnF d'un enseignement de 14 semaines sur le thème « Internet en campagne : les usages des archives électorales de la BnF », auquel participera un groupe d'une quinzaine d'étudiants inscrits en Master à Sciences Po.

- **un volet scientifique** : une équipe de chercheurs en science politique et en sociologie des médias encadrera le travail des étudiants au sein du pôle associé et à la BnF et conduira des recherches complémentaires sur le même corpus d'archives électorales. Les grandes orientations de recherche retenues concernent l'appropriation des technologies de l'Internet par les formations politiques et les internautes de la société civile ainsi que l'émergence de ce nouveau type de sources dans les pratiques de recherche universitaire en sciences humaines.

- **un volet documentaire** : au cours de cette expérimentation, la BnF testera plusieurs applications informatiques utiles à la recherche et à la navigation dans les archives d'Internet. Elle observera également les comportements et les réactions des étudiants et des chercheurs afin d'analyser les situations de service et de médiation futurs.

**Sciences Po** est responsable des volets **pédagogique et scientifique** du projet : sélection des chercheurs associés au projet, suivi et valorisation de leurs activités.

**La BnF** est responsable du **volet documentaire** du projet. Elle mettra à disposition des étudiants et des chercheurs les collections d'archives électorales et les installations techniques nécessaires à la conduite de cette expérimentation dans ses salles de lecture. Elle apportera son expertise et son concours afin de faciliter l'utilisation de ce dispositif par les utilisateurs.

## Glossaire

### Baladodiffusion / *Podcasting*

Pratique qui consiste à diffuser, en général gratuitement, des enregistrements (des *podcasts*) audios ou vidéos sur Internet. La baladodiffusion permet aux internautes de télécharger ces enregistrements, souvent automatiquement, via des abonnements à des flux RSS ou Atom, sur le disque dur de leur ordinateur personnel pour une écoute immédiate ou différée sur leur baladeur numérique.

**Blogosphère politique** : l'ensemble des blogues politiques existants ou la communauté des rédacteurs de ces blogues.

### Mots clés (achat de)

Pratique publicitaire par laquelle un annonceur peut s'assurer d'une position favorable dans les pages de résultats des moteurs de recherche en achetant et/ou en enchérissant sur les mots clés correspondant aux requêtes des internautes sur lesquelles il souhaite être présent. Les messages publicitaires ainsi publiés sont affichés par les moteurs de recherche dans les zones de liens commerciaux (appelés aussi liens promotionnels ou sponsorisés) en marge des résultats ou sur des sites partenaires. Ils permettent d'augmenter la notoriété du site des annonceurs.

### Pistage / *Tracking*

Procédé qui consiste à suivre et à mémoriser le parcours d'un internaute sur un site. Il permet, notamment, de rémunérer les sites partenaires qui envoient des clients et de proposer des services ou des annonces publicitaires en fonction du profil du visiteur.

### Publipostage / *E-mailing*

Envoi en nombre d'un courrier électronique à des internautes. Le publipostage est utilisé le plus souvent dans le cadre d'une campagne de promotion d'un service ou d'un produit. Dans le cadre de la communication politique, il est souvent informatif et encourage les nouvelles adhésions.

### SPAM, *spamming*

Qualificatif péjoratif pour un courrier électronique publicitaire envoyé à de nombreux internautes, ou simplement un courrier non sollicité. *SPAM* est l'acronyme de *Spiced Pork And Meat* (Pâté de viande épicé à base de porc et de viandes diverses).

Les SPAM vantent le plus souvent les mérites de services pornographiques, de médicaments de type dopage sexuel, de logiciels, ou proposent des investissements lucratifs. Ils encombrant les boîtes aux lettres des internautes.

Le terme de *spamming* est utilisé pour désigner l'envoi en nombre de SPAM.

### Taux de clic

Pourcentage utilisé pour mesurer le succès de campagnes publicitaires de type *spamming* ou achat de mots clés.

### Web-TV

Système de diffusion permettant aux internautes de regarder une chaîne de télévision en ligne. Les contenus peuvent être diffusés en temps réel ou en différé, de manière continue ou sous la forme d'émissions.

### Wiki

Logiciel spécialisé qui rend les pages d'un site Internet librement modifiables par tout visiteur. Les wikis sont utilisés pour faciliter l'écriture collaborative de documents. Le wiki le plus connu est l'encyclopédie libre, universelle et multilingue Wikipédia (<http://fr.wikipedia.org>).