

# Le recours aux SIC dans la mise en place du dépôt légal du web à l'INA

par AUBERT Aurélie, « [aurel.aubert@wanadoo.fr](mailto:aurel.aubert@wanadoo.fr) »

CIM (Communication Information Médias) EA 1484 - Paris 3 Sorbonne Nouvelle

Cette contribution propose le récit d'une démarche expérimentale menée en collaboration avec l'INA sur le projet de dépôt légal du web, actuellement en cours d'élaboration. L'expertise de chercheurs en SIC a été sollicitée dans la phase de consultation des sites archivés, consacrés à la présidentielle de 2007, sites qui ont été collectés à cette époque, par l'INA. Notre contribution s'est faite dans le cadre d'une recherche menée avec le laboratoire Communication et Politique du CNRS sur les rôles et usages des médias pendant la présidentielle.

La consultation des archives du web par le chercheur en SIC soumet plusieurs questions au coeur de la discipline, car une réflexion approfondie sur la notion de "dépôt légal" permet d'articuler les notions d'espace public, d'identité mémorielle et d'innovations technologiques.

Nous revenons sur les questions relatives au "genre" des sites internet archivés et à celles portant sur l'élaboration des notices descriptives devant intégrer, à terme, un certain nombre d'éléments propres aux sites web : dynamisme, fréquence de mise à jour, auteurs, caractérisation des contenus... Ces éléments posent la question de la terminologie et de la typologie à mettre en oeuvre pour rendre compte de la richesse d'une archive.

Le récit de cette expérience fait apparaître les synergies envisageables entre les chercheurs en SIC spécialisés sur les médias et ceux travaillant sur la documentation, notamment au travers de la recreation de l'interactivité du web, au sein des archives.

*Mots-clés* : Dépôt légal du web - Archives - Interactivité - Mémoire - Elections présidentielles

This article explains our experimental test of the INA consultation devices of web archives in France, as a researcher specialized in Information and Communication. The analysis about the notion of "registration of copyright of the Internet" deals with questions such as public sphere, memorial identity and technological innovations. This experience shows that researchers have to cast new light on terminology and typology of web sites.

*Keywords* : Registration of copyright of the Internet - Archives - Interactivity - Memory - Presidential elections

## Le recours aux SIC dans la mise en place du dépôt légal du web à l'INA<sup>[1]</sup>

La production numérique sur internet est sur le point d'être conservée, en France, grâce à l'une des dispositions de la loi sur le droit d'auteur et les droits voisins (dadvsi) qui stipule la mise en place d'un dépôt légal pour les sites internet. La tâche est confiée conjointement à la BNF (Bibliothèque nationale de France) et à l'INA (Institut national de l'Audiovisuel), chargés de

constituer la mémoire numérique nationale. Le périmètre de l'INA concerne les sites liés à la communication audiovisuelle et aux industries culturelles. L'INA sera donc en charge de l'archivage des flux et devra sélectionner, collecter et conserver les sons et images animées du web national.

Le procédé d'archivage du web est, depuis plusieurs années, et bien avant la promulgation de la loi du 1<sup>er</sup> août 2006, en cours d'expérimentation à l'INA comme à la BNF, afin de déterminer d'une part, les moyens à mettre en œuvre pour la captation des sites, d'autre part le périmètre de cette captation. Cette réflexion s'articule autour de différentes notions qui mettent en jeu l'expertise de plusieurs types d'acteurs : archivistes, documentalistes, ingénieurs informaticiens et chercheurs (notamment en sciences de l'information et de la communication) destinés à utiliser ces contenus archivés. Une phase de test du dispositif de consultation de ces archives a été lancée au printemps 2007 sur les sites web dédiés à la présidentielle française. L'INA a, pour cette occasion, archivé quelques 2200 sites consacrés à cette élection entre janvier et juin 2007. Notre contribution à ce projet de « dépôt légal du web » a eu lieu pendant la phase de consultation des archives ainsi collectées. En raison de notre participation à un projet de recherche sur l'étude du rôle et des usages des médias pendant la dernière campagne électorale, notre expertise a été sollicitée.

Cet article propose le récit d'une démarche expérimentale menée en collaboration avec l'Institut national de l'Audiovisuel en vue de la conservation du patrimoine numérique. Cette expérience se situe au cœur de la discipline des SIC car elle touche, à la fois, à la notion d'espace public, à celle d'identité et aux innovations technologiques. Nous présentons ici le point de vue du chercheur en SIC que nous sommes quant aux enjeux liés à l'archivage du web. Cependant, notre propos reviendra également sur la démarche des documentalistes dans cette entreprise dans la mesure où ces derniers sont à l'origine de l'élaboration et de la conception des procédés d'archivage.

## **Le principe du « dépôt légal » : quels enjeux pour le web ?**

### **Définition du dépôt légal et périmètre d'action**

Le « dépôt légal » rassemble tous les documents français destinés à une diffusion publique. Par la conservation systématique d'au moins un exemplaire de tout document publié, il participe à la création de la mémoire de la nation. Ce principe s'inscrit dans la continuité des missions de conservation de la BNF et de l'INA, dans la mesure où le dépôt légal est l'obligation faite par la loi à tout éditeur, imprimeur, producteur, distributeur, importateur de documents d'effectuer un dépôt auprès des organismes désignés par la loi.

Initialement promulgué pour les livres imprimés et les périodiques, le « dépôt légal » s'est, au cours de l'Histoire, étendu à tous les types d'expression et de création, en intégrant les nouvelles techniques, au fur et à mesure de leur apparition : les imprimés sont ainsi conservés depuis 1537, les estampes (dont les cartes et les plans) depuis 1648, les photographies depuis 1925, les vidéogrammes depuis 1975, les contenus audiovisuels depuis 1992. La loi du 1<sup>er</sup> août 2006 a étendu le champ du dépôt légal aux « signes, signaux, écrits, images, sons ou messages de toute nature qui font l'objet d'une communication au public par voie électronique ».

La banalisation de l'usage d'internet qui se note, par exemple, dans l'abondance de références à des liens hypertextes dans les productions universitaires, intellectuelles ou éditoriales a conduit à envisager cette conservation aujourd'hui entendue comme légitime. Celle-ci se révèle nécessaire pour une société de l'information encline à maîtriser ses sources. Les enjeux du dépôt légal du web dépassent le simple cadre de la loi et s'envisagent aussi sur un plan philosophique et économique, comme le pense, par exemple, Louise Merzeau (2003) pour qui :

« Le projet de dépôt légal d'Internet introduit [...] une double courbure dans l'économie temporelle du web. D'une part, il fait passer dans le champ de la stratégie des processus qui relevaient de la tactique, en convertissant un espace indéfini - celui des trajectoires imprévisibles et fugaces de l'information - en un lieu technocratiquement bâti, circonscrit et géré. D'autre part, il offre une alternative aux stratégies de traçabilité [...] en apportant la garantie de l'institution à une rétention jusqu'alors soumise aux seules lois de l'innovation technologique et de la concurrence ».

Plusieurs expériences différentes ont été tentées pour penser le principe de l'archivage de contenus aussi disparates et évolutifs que les sites internet, aujourd'hui complétés par les blogs, portails, etc... La BNF a, par exemple, expérimenté le principe d'une « photographie » de la toile à un moment donné. Elle a, parallèlement, effectué des collectes thématiques concernant des événements particuliers : ce fut le cas pour les sites électoraux de 2002 (élections présidentielles) et de 2004 (élections régionales et européennes). La sélection thématique des sites électoraux de 2002 et 2004 a ainsi généré la collecte de 3500 sites tandis que la photo instantanée du domaine .fr réalisée en début d'année 2005 représente une collecte de 118 millions de fichiers[2].

Cette nouvelle disposition de la loi pose, en partie, problème aux archivistes en raison de la nature même du web qui est un « hypermédia » : un site web est, en effet, un ensemble de pages web liées sémantiquement. Il contient souvent du texte, des images, parfois de la vidéo, des fichiers en *streaming*, des fichiers en téléchargement. Se pose donc la question de son mode d'archivage, de sa collecte, mais aussi de sa consultation puisque au delà de la conservation des données, il est nécessaire de travailler sur les moyens d'une consultation aisée et optimale pour les chercheurs. Il faut donc indexer et ordonner le contenu, puis créer un moteur de recherche qui permette aux utilisateurs de s'y retrouver aisément.

Mehdi Garsallah (2004) rappelle les trois écoles qui ont émergé au moment des premières réflexions sur la sauvegarde du patrimoine numérique en ligne. On dénombre :

- D'une part les projets qui ne visaient uniquement que l'archivage des documents en ligne, c'est à dire la plupart des pages html ou au format de texte (Microsoft word) ou encore des fichiers Adobe Acrobat au format Pdf.
- D'autre part, les équipes qui privilégiaient le web comme une création à part entière mais proposaient un archivage statique des pages en les passant, par exemple, sur microfilm

- Enfin, les projets prenant en compte la nature hypermédiasique et dynamique du web qui proposaient, avec toutes les difficultés que cela comporte, un archivage le plus proche possible de ce qui est réellement en ligne.

C'est cette troisième école qui est privilégiée par l'INA, mais les contraintes techniques d'internet ainsi que les contenus hétérogènes auquel l'archiviste est confronté ouvrent des pistes de réflexion complexes. Ainsi, l'archive ne doit pas représenter un simple amas de pages collectées, sans indexation, mais doit proposer un ensemble de pages cohérentes en réponse à une requête du chercheur. En outre, les techniques évoluent en un laps de temps très rapide, évolution qui ne doit pas être un handicap à une consultation future. Enfin, le web est construit sur une logique d'interaction : la collecte d'un site internet doit donc reproduire cette interaction avec l'utilisateur afin que la consultation du document web sous forme d'archive reproduise la consultation qui en est faite dans sa structure et dans sa publication.

## **De l'identification des sites à leur consultation : les étapes du projet de « dépôt légal du web »[\[3\]](#)**

Les archivistes et ingénieurs de l'INA ont considéré que c'est la structure de l'archive qui doit orchestrer la consultation des sites dans le dispositif : « les difficultés imposées par le médium web apparaissent donc essentiellement liées à des questions de délimitation, à la fois spatiales et temporelles », rappelle Thomas Drugeon, de l'INA (Drugeon, 2005 : 60). Ces questions ont été au cœur de la conception et de la mise en place du dépôt légal du web par cette institution qui a donc élaboré plusieurs étapes : l'identification des sites du domaine, leur collecte, leur indexation, leur stockage, puis leur consultation. C'est à l'étape de la consultation que notre compétence de chercheur a été sollicitée.

La phase d'identification consiste d'abord à déterminer le périmètre d'archivage et les sites concernés par ce procédé (selon Jean-Michel Rhodes, directeur de l'inathèque « un site français est un site dont l'adresse postale du propriétaire est sur le territoire. Cela représente environ 300 000 sites. 150 000 en .fr et à peu près autant en .com, .org et .net »[\[4\]](#)). D'autre part, cette phase consiste à décider la fréquence à laquelle un site doit être collecté, selon son dynamisme, à savoir la fréquence de mise à jour des pages.

Les instruments de collecte, chargés de reproduire l'interaction entre l'interface hypermedia qu'est le site web et l'utilisateur, et de restituer la structure d'un site sont des robots « crawlers », leur fonctionnement est expliqué par Thomas Drugeon (2005 : 58) :

« Le principe en est simple : le robot rapatrie une page d'un site, l'archive répertorie l'ensemble des hyperliens qu'elle contient, puis rapatrie chacune des pages pointées par ces hyperliens, et procède de la même manière pour chacune d'entre elles et ce de manière itérative, jusqu'à ce que l'ensemble des liens aient été suivis ».

Le résultat de la collecte est ensuite indexé c'est à dire que l'archive est documentée (on en précise la nature, la date par exemple) et éventuellement enrichie (connexion avec d'autres sites du même type, cartographie du web). Le résultat de la phase de collecte est enfin stocké pour être consulté par des chercheurs sur une interface visant à recréer l'interactivité du web archivé. Le principe de la consultation par le chercheur consiste à se déplacer dans l'archive selon un procédé temporel, comme s'il naviguait dans le temps : les choses se déroulent donc comme s'il consultait réellement le web, à la date qu'il a choisie.

Dans sa démarche de mise au point de l'outil de consultation des archives du web, l'INA s'inscrit dans le même processus que ce qui avait été envisagé au moment de la mise en place du dépôt légal de la télévision avec l'élaboration de l'outil Media Corpus. Jean-Claude Soulages, à l'origine, notamment, avec les membres du CAD (Centre d'Analyse du Discours) de cette synergie entre l'INA et les chercheurs rappelle l'intérêt d'un outil comme Mediacorpus :

« Avec la création de l'Inathèque, les bases documentaires professionnelles de l'Institut national de l'Audiovisuel, constituées à partir du travail d'indexation systématique des documents audiovisuels, ont été reformatées en vue d'une démarche de type scientifique. Le prototype d'interface conçu et élaboré avec le concours du CAD - aujourd'hui mis à la disposition des chercheurs sous sa forme évoluée (Mediacorpus) - permet de circonscrire et d'autonomiser de vastes corpus de documents à partir de critères discriminants » (Soulages, 2003).

Tout comme la création des dispositifs de consultation des archives de contenus télévisuels a été pensée dans une démarche scientifique pour répondre aux besoins caractéristiques de chercheurs, celui du web doit être conçu dans les mêmes termes, en fonction également des contraintes et dispositifs inhérents à internet, support spécifique et hétérogène.

## **Le web électoral : un premier test**

Une première phase de test de l'outil de consultation a été réalisée au printemps 2007 sur un corpus de sites, blogs et portails consacrés aux élections présidentielles. L'INA a travaillé, pour cette opération avec RTGI (Réseaux, Territoires & Géographie de l'information), jeune société de « veille informationnelle ». Le périmètre du corpus à archiver s'est fondé sur la typologie et le recensement proposé par le site [www.bonvote.com](http://www.bonvote.com) offrant une sélection de quelque 1400 sites et blogs relatifs à cet événement. L'exhaustivité n'étant pas totale, d'autres prospections, réalisées par RTGI, ont été à l'étude pendant la phase d'archivage. Cette campagne de prospection a eu pour but d'étendre le corpus. Le classement des sites ainsi collectés s'est fait selon une organisation topologique ou sémantique, en fonction également du nombre de liens hypertextes ainsi que du nombre de mises à jour. L'originalité de cette opération a consisté en la « fréquence de rafraîchissement des pages, jusqu'à quatre fois par jour, les jours précédant le vote »<sup>[5]</sup>.

Pendant cette phase d'archivage, possibilité m'a été donnée de consulter ces archives (navigation temporelle en fonction de la date). Le dispositif n'était doté, à l'époque, d'aucune notice relative à chaque site (sur le mode des notices des émissions télévisuelles mises au points par les documentalistes de l'INA). En revanche une recherche par « mots-clefs » permettait d'effectuer une recherche dans la page.

## **La confrontation de deux logiques : celle du chercheur, celle du documentaliste**

### ***Un projet collectif : rôle et usages des médias lors de la campagne présidentielle de 2007***

Au printemps 2007, nous avons participé à un projet de recherche collectif sous la houlette du laboratoire Communication et Politique du CNRS (LCP), consistant à interroger les rôles et les usages des médias lors de la présidentielle. Ce projet de recherche a réuni une vingtaine d'enseignants chercheurs étudiant la médiatisation de l'ensemble de la campagne présidentielle sur différents supports, selon trois axes qui avaient été définis préalablement, à savoir :

- Les thèmes de campagne et leur mise en agenda
- La fabrication de l'information
- La réception de l'information

Dans cette recherche, notre propre contribution s'est axée sur les nouveaux supports communicationnels générés par le web 2.0 (blogs, sites politiques ainsi que les sites d'information « participatifs »), dans le but d'étudier de manière approfondie les enjeux d'internet en temps de campagne électorale. Les présupposés de notre démarche étaient les suivants : il s'agissait, tout d'abord, de réintroduire le rôle des publics et des citoyens dans la couverture de la campagne et la médiatisation de certains enjeux électoraux. Parallèlement, nous avons cherché à réinterroger la croyance collective en la potentialité d'internet à structurer le rapport des citoyens au politique et à influencer le résultat des élections depuis le référendum de 2005 où des sites personnels avaient contrebalancé le poids de la parole des médias traditionnels (plutôt orienté pour le oui). L'une des questions centrales de cette recherche a consisté à se demander, notamment, si les citoyens internautes ont proposé une couverture alternative de la campagne électorale. Une étude de cas a été proposée avec l'analyse du site de « journalisme participatif » le plus abouti au moment de la campagne : Agoravox<sup>[6]</sup>.

La possibilité d'avoir accès à l'ensemble du web présidentiel a cependant contribué à une réflexion sur des questions plus vastes, questions auxquelles nous n'avons pas pu, pour l'instant, totalement répondre. Nous les livrons cependant afin d'envisager les possibilités en terme de

recherche, liées à cette interface de consultation. Ainsi, la possibilité de naviguer temporellement dans l'archive permet d'évaluer le dynamisme des sites (leur remise à jour, le nombre de commentaires suscités par un *post*, par exemple). Sur un autre registre, dans la perspective d'une analyse comparative entre médias traditionnels et nouveaux médias, on peut également évaluer le mode opératoire des blogs vis-à-vis des médias traditionnels (citation des informations de la télévision, usage des vidéos détournées, utilisation des informations données par la presse...). Notre démarche a donc combiné une approche quantitative et une approche qualitative afin de rendre compte au mieux de la couverture de la campagne sur le web.

## ***Les attentes de l'INA***

Les objectifs des acteurs de l'INA n'étaient pas de même nature : leur principal souci consistant à comprendre les logiques des chercheurs travaillant sur les différents supports de l'internet afin de faire évoluer la maniabilité de l'outil de recherche. Le chercheur est alors vu comme un évaluateur de l'interface de consultation concernant, notamment, les problèmes d'archivage afin de signaler les éventuelles difficultés techniques. L'expertise est ensuite sollicitée pour renseigner, compléter ou modifier les champs d'application de l'outil de consultation.

## ***Une démarche expérimentale***

Après avoir pris en main et testé l'outil de consultation, plusieurs questions se posent au chercheur en SIC. Notre examen s'est effectué sur une sélection de sites à caractère informatif ou politique, selon une navigation temporelle (l'interface propose la « dernière version enregistrée », la « version actuellement en ligne » ainsi qu'une collection de dates auxquelles le site a été réactualisé).

La première difficulté qui apparaît au chercheur devant la multiplicité des supports est leur laborieuse classification. Les informations fournies par le site d'un parti, celles que l'on trouve sur un portail d'information ou encore celles qui sont signées par un blogueur plus ou moins reconnu de la blogosphère ne peuvent, à l'évidence, être exploitées de la même manière dans la perspective qui fut la nôtre, à savoir évaluer la couverture d'une campagne électorale. La première interrogation qui a été soulevée dans les discussions préliminaires avec les ingénieurs et documentalistes de l'INA a donc porté sur la notion de « genre » des sites internet. Autant le renseignement de ce champ semble aisé pour les documentalistes de l'INA chargés d'indexer des programmes télévisés, autant il apparaît dans toute sa complexité en ce qui concerne le web. Cette difficulté tient, en partie, à l'hétérogénéité des contenus des sites qui peuvent inclure aussi bien des chats, des forums de discussion, des posts, des pages html... La notion de producteur - la personne morale à l'origine de la création du site ou de la page - est également problématique.

Ces difficultés conduisent à s'interroger sur la notion de « notice » relative à un site internet et aux informations que cette notice doit pouvoir, à terme, contenir. C'est toute la terminologie du

web qui apparaît ici au cœur des problématiques de classification. Dans l'esprit des concepteurs des archives web de l'inathèque, la définition d'une « notice de collection » correspondrait à un ensemble de pages dont on décrirait, de la manière la plus stable possible, la structure et l'« intention », ainsi que les services proposés (service primaire, service secondaire). Il a ainsi été évoqué la possibilité, pour les chercheurs, à terme, de remplir eux-mêmes les notices en décrivant un corpus, et en proposant une classification. Les descripteurs personnalisés posent cependant la question de la stabilisation des normes, loin d'être effectuée.

En outre, les différents enjeux de l'élaboration d'une classification sont apparus par rapport à la présentation hiérarchique des sites archivés proposée. En effet, ceux-ci étaient présentés dans une fenêtre, selon une classification reposant, entre autres, sur la couleur politique. Dans cette perspective, les sites informationnels ou les blogs de commentaires de l'actualité trouvent difficilement leur place : les enjeux d'une classification par genre, par « intention » ou par type de contenus apparaissent donc dans toute leur complexité.

Cette problématique rejoint enfin celle de l'élaboration d'un thésaurus : dans Mediacorpus, la recherche par mot clef est possible grâce à un glossaire les regroupant. Ce mode de recherche est, pour l'instant, impossible sur les archives du web. En l'état actuel des choses, aucune recherche thématique ne peut être faite sur l'ensemble du corpus, indépendamment d'une recherche menée sur le nom du site. Ainsi, si l'on souhaite rechercher les sites proches de l'UDF au moment de la dernière présidentielle, seuls ceux portant ces trois lettres dans leur nom de domaine seront proposés. À terme, c'est une navigation par mots-clefs ou par voisinage sémantique qui est envisagée, elle reste à mettre en place. L'enrichissement de l'indexation est également à l'étude avec la possibilité, notamment, de cartographier les liens de références ou les citations faites d'un site par un autre : ces données permettent de proposer une carte des communautés en ligne. Cet aspect est travaillé par la société RTGI<sup>[7]</sup>.

Selon les perspectives de recherche qui furent les nôtres à ce moment donné, il nous a semblé essentiel de pouvoir intégrer à l'outil de consultation un système de book mark ou marque page afin de renseigner plusieurs indications : le degré de mise à jour de tous ces sites ; la fréquence et le volume des commentaires (entre 0 et 5, entre 5 et 10, plus de 10, plus de 30...) : ce type d'indications revêt en effet une extrême importance pour évaluer le dynamisme d'un site. De même, il était nécessaire de pouvoir examiner, de manière qualitative, la campagne menée par les internautes sur ces supports.

Notre démarche a ainsi consisté à noter, de manière récurrente, sur un certain nombre de sites à caractère politique, le genre d'information proposées à la lecture : actualité du ou de la candidate, entreprise de déstabilisation du camp adverse, billets d'humeurs liés à l'actualité... autant de genres différents pour des posts ou des contenus qui nécessitent une réflexion quant à la terminologie d'une part et la typologie à mettre en place d'autre part.

Reste enfin la question fondamentale du rapport entre internet et les médias dits « traditionnels ». La réutilisation de dépêches, d'articles parus sur les supports interactifs de quotidiens ou d'hebdomadaires, la diffusion de contenus vidéos d'émission télévisées forment l'armature de la plupart des blogs. L'internet se nourrit abondamment de ce qui est diffusé à la télévision. Nous avons pu constater une utilisation intensive des débats télévisés et extraits d'émission dont on trouve de nombreuses traces. La perspective de recherche consistant à évaluer les interactions entre internet et les médias traditionnels suppose donc de prendre en compte quantitativement et qualitativement ces interactions et de pouvoir intégrer, à terme, dans une notice, ce type d'information en précisant, par exemple, si la citation du document audiovisuel est totale, partielle, tronquée, détournée, etc. C'est une des remarques qui a donc été formulée aux documentalistes.

A l'heure actuelle, alors qu'un examen complémentaire reste à faire, nos propres conclusions sur la couverture de la campagne électorale font apparaître un faible niveau d'interaction sur les blogs de particuliers : les internautes sembleraient être allés, de préférence, sur les sites des grands journaux qui ont renouvelé leur offre concernant les élections présidentielles. Nous avons également montré que la couverture de la campagne proposée par le site participatif Agoravox était loin d'être véritablement alternative à celle de la télévision : l'espace dévolu aux petits candidats, par exemple, fut moins important que sur France 2, contrainte par le CSA. La blogosphère semble avoir manqué le tournant de cette élection pour se révéler réelle pourvoyeuse d'informations inédites. Les éléments quantitatifs et qualitatifs permettant d'étayer ces affirmations ont pu être collectés grâce à l'archivage des sites web : reste à mettre en place un outil de consultation aussi maniable pour le chercheur que Mediacorpus.

La collaboration entre les documentalistes et archivistes de l'INA et des chercheurs en SIC plutôt axés sur la question des médias (ici nous ne représentons pas la branche « documentation » des SIC) montre les possibles synergies entre ces deux axes de recherche dont les concepts et méthodes doivent être travaillés de concert en vue de la mise en place de ce projet de « dépôt légal ». Notre propre réflexion porte sur les enjeux d'internet par rapport à l'ouverture de l'espace public et au rôle du citoyen dans l'élaboration de l'information : elle s'articule autour de la nécessité de comprendre les modes d'appropriation des techniques par les citoyens, la nature de leurs publications sur internet et leur mode de consultation, a posteriori. La recreation de l'interactivité du web, au sein des archives, souligne, par exemple, la rencontre entre ces différents enjeux.

La logique des documentalistes visant à élaborer une base de données complète et interactive s'accomplit en partie en interaction avec des chercheurs qui peuvent orienter la manière dont seront conçus les outils de recherche. En effet, la classification, la mise au point de mots-clefs descriptifs pour faciliter la consultation ou la rédaction de notices normées sont autant d'enjeux relatifs aux connaissances et aux savoirs qu'une société se construit.

# Bibliographie et ressources électroniques

Aubert A., 2009, « La couverture des présidentielles par les sites de journalisme amateur », Actes du colloque Médias et élections : analyses interdisciplinaires de la campagne présidentielle 2007, INA/ laboratoire LCP, (à paraître).

Bimber B. et al., 2003, Campaigning on line. The Internet in US elections, Oxford ; New York, Oxford University press, 224 p.

Greffet F., 2001, « [Les partis politiques français sur le Web](#) » in Andolfatto, A. Greffet, F. Olivier, L. (dir) Les partis politiques : quelles perspectives ? Paris, L'Harmattan (coll. Logiques politiques), 267 p.

Drugeon T., 2005, « L'INA et la mission de dépôt légal », Migrations n°23, pp. 56-70.

Fouetillou G., Ecologie de la blogopole, in Observatoire Présidentielle [en ligne], 2006.

Disponible sur <http://www.observatoire-presidentielle.fr/?pageid=31> [consulté le 17 janvier 2007].

Cardon D. Le design de la visibilité, un essai de typologie du web 2.0, in : Internet Actu [en ligne], 1er février 2008, [consulté le 10 avril 2008]. Disponible sur : <http://www.internetactu.net/2008/02/01/le-design-de-la-visibilite-un-essai-de-typologie-du-web-20/>

Gharsallah M., Dépôt légal des publications électroniques et préservation patrimoniale du web français, in : Archivesic [en ligne], décembre 2004, [consulté le 18 mars 2008]. Disponible sur : [http://archivesic.ccsd.cnrs.fr/sic\\_00001311.html](http://archivesic.ccsd.cnrs.fr/sic_00001311.html).

Merzeau L., 2003, « Web en stock, le dépôt légal d'internet », in : Les cahiers de médiologie n°16 Éternel éphémère. Disponible sur : <http://www.merzeau.net/> [consulté le 20 juin 2007].

Rallet A., 2006, « Technologies de l'information et de la communication. Un enjeu économique de société » pp. 145-160 in Olivési S. (dir.), Sciences de l'information et de la communication, Objets, savoirs, discipline, Grenoble, PUG, 286 p.

Rébillard F., 2007, Le web 2.0 en perspective, Paris, l'Harmattan (coll. Questions contemporaines), 162 p.

Soulages J.C., « La thématique Europe à l'intérieur des journaux télévisés français (1951-2003) », in : Mundaneum [en ligne], 2003. Disponible sur <http://www.mundaneum.be/fichiers/pdf/soulages.pdf> [consulté le 12 mars 2008].

Vedel T., 2003, « Internet et les pratiques politiques » pp. 189-214 in Gingras A.M (dir.) La communication politique, états des savoirs, enjeux et perspectives, Saint Foy, Presses de l'université du Québec, 295 p.

---

[1] La rédaction de cet article n'aurait pu se faire sans les échanges menés avec les différents acteurs de l'INA qui travaillent sur le projet de dépôt légal du web. Thomas Drugeon, Nicolas Delaforge et Renaud Huerta m'ont apporté leur expertise technique ; Christine Barbier Bouvet et Denis Maréchal ont, notamment, joué un rôle important dans la collaboration entre l'inathèque et les chercheurs. Je les remercie sincèrement.

[2] Selon Catherine Lupovici directrice du département de la bibliothèque numérique à la BNF, citée sur ZDNet.fr le 14 décembre 2005 : <http://www.zdnet.fr/actualites/internet/0,39020774,39295404,00.htm> [consultation le 10 avril 2008].

[3] Cette partie technique est expliquée par Thomas Drugeon dans son article « L'INA et la mission de dépôt légal » in *Migrance* n°23, premier semestre 2005, pp. 56-70 auquel nous nous référons principalement ici.

[4] Cité par Medhi Garsallah (2004).

[5] Selon Jean-Michel Rhodes, in Astrid Girardeau « Le web, nouvelle terre d'élections », *Libération* du 14 mai 2007.

[6] C'est cette étude de cas qui a été présentée au colloque rendant compte des résultats de cette recherche collective les 10 et 11 janvier 2008 à l'INA. Aurélie Aubert, « La campagne électorale de 2007 sur les blogs et les sites de "journalisme amateur" », communication au colloque Médias et élection : analyses interdisciplinaires de la campagne présidentielle 2007, INA/ laboratoire LCP, 10-11 janvier 2008.

[7] Un exemple de cartographie des communautés du web pendant la présidentielle est disponible sur le site <http://www.observatoire-presidentielle.fr/> : on peut y consulter la blogopole, mise en scène de la cosmogonie évolutive des blogs, rangés selon leur sensibilité politique et reliés les uns aux autres en fonction de leurs influences réciproques.