

Diplôme de conservateur des bibliothèques

Mémoire de DCB / décembre 2008

**Définir une politique d'archivage du  
web régional en bibliothèque  
municipale.  
L'exemple du web forézien.**

**Bruno BONNENFANT**

Sous la direction de M. Thierry Veyron

Conservateur responsable du Patrimoine et du Fonds régional  
à la Médiathèque de Saint-Etienne



## **Remerciements**

*Je tiens à remercier Thierry Veyron, qui a proposé ce sujet et suivi ma recherche, et à exprimer toute ma gratitude à Gildas Illien, Frédérique Joannic-Seta et Clément Oury qui m'ont accueilli avec chaleur à la BnF et ont mis à ma disposition la documentation relative à la collecte du web électoral.*

*Je tiens également à adresser tous mes remerciements aux professionnels qui ont accepté de répondre à mon enquête, et particulièrement à François Berquet (Troyes), Florence Courtial (Toulouse), Christelle Guillet (Rouen), Bernard Huchet (Caen), Isabelle Lang (Aix-en-Provence), Anne Meyer (Lyon), Françoise Michelizza (Nice), Olivier Morand (Orléans).*



## **Résumé :**

*La volonté d'archiver le web local en France se heurte aux dispositions de la loi de 2006. Cette étude examine les conditions d'un partenariat entre les bibliothèques municipales et la bibliothèque nationale. En outre, elle propose une approche méthodologique qui tient compte de la spécificité du document web et elle s'appuie sur un exemple régional : celui du Forez à la bibliothèque de Saint-Etienne.*

## *Descripteurs :*

*Archives de l'Internet - - Forez(France)*

*Bibliothèques publiques - - Développement des collections*

## **Abstract :**

*The willingness to archive local web sites in public libraries is running up against the provisions of the 2006 Law. This study examines the conditions of a partnership between public libraries and the national library. In addition, it proposes a methodological approach that takes into account the specificity of the Web document and it relies on a regional example : the Forez at the Saint-Etienne library.*

## *Keywords :*

*Web archives - - Forez(France)*

*Public libraries - - Collection development*

## **Droits d'auteurs**

Droits d'auteur réservés.

Toute reproduction sans accord exprès de l'auteur à des fins autres que strictement personnelles est prohibée.
--



# Table des matières

INTRODUCTION .....	9
<b>INTERET DE L'ARCHIVAGE DU WEB EN BIBLIOTHEQUE MUNICIPALE. ETAT DES PRATIQUES.....</b>	<b>11</b>
1. POURQUOI ARCHIVER LE WEB REGIONAL ?.....	11
1.1 <i>Quelques statistiques sur la consultation de l'internet</i> .....	12
La consultation de l'internet en France.....	12
Le domaine français .....	12
Une étude des usages dans les régions françaises .....	13
1.2 <i>Qu'entend-on par « web régional » ?</i> .....	15
Essai de définition : l'objet et la méthode.....	15
Une démarche consciente de constitution d'une collection.....	16
Un objet d'étude autonome .....	16
1.3 <i>Intérêt du web régional en bibliothèque municipale</i> .....	17
Enrichissement et continuité des collections.....	17
Valorisation de la bibliothèque .....	17
2. LE CADRE JURIDIQUE ACTUEL .....	18
2.1 <i>Le dépôt légal des sites</i> .....	18
2.2 <i>Propriété intellectuelle et droit moral : la consultation des archives</i> .....	19
2.3 <i>Quel décret d'application ?</i> .....	20
3. LA BNF EN CAMPAGNE : RETOUR SUR UNE EXPERIENCE DE COOPERATION AVEC LES BIBLIOTHEQUES .....	20
3.1 <i>Une démarche patrimoniale et expérimentale</i> .....	20
De la collecte large à la collecte thématique.....	20
La campagne électorale, objet patrimonial et champ d'expérimentation .....	21
3.2 <i>Méthodologie : une entreprise associant les BDLI</i> .....	22
La prise en compte de l'échelle locale.....	22
Les BDLI : partenariat et accompagnement .....	23
La tâche des sélectionneurs.....	23
Eléments de bilan .....	24
3.3 <i>Intérêt de la coopération</i> .....	25
Des bénéficiaires réciproques.....	25
Valorisation de l'expérience au sein des BDLI .....	25
Perspectives .....	26
4. ELABORATION D'UN SCENARIO POSSIBLE POUR UN RESEAU D'ARCHIVAGE DU WEB ..	26
4.1 <i>Une structure partenariale</i> .....	26
Une collecte partagée .....	26
La BDLI, échelon intermédiaire du réseau .....	27
4.2 <i>Quels projets communs ?</i> .....	28
Exhaustivité ou sélectivité ?.....	28
Quelques pistes de projets.....	28
4.3 <i>La place des bibliothèques municipales</i> .....	29
Le soutien des agences de coopération ? .....	29
Quel bénéfice pour les bibliothèques municipales ?.....	30

<b>METHODOLOGIE POUR UNE POLITIQUE D'ARCHIVAGE.....</b>	<b>31</b>
1. INTEGRER L'ARCHIVAGE DU WEB A LA POLITIQUE DOCUMENTAIRE DE L'ETABLISSEMENT .....	31
1.1 <i>Le fonds local en bibliothèque municipale</i> .....	31
Un domaine de référence .....	32
Patrimonialité du fonds local .....	33
Territorialité .....	34
1.2 <i>La politique documentaire à l'heure de l'internet</i> .....	35
1.3 <i>Publics</i> .....	37
2. PARAMETRES A PRENDRE EN COMPTE.....	39
2.1 <i>Quelle unité de collecte ?</i> .....	39
2.2 <i>Le critère de profondeur</i> .....	40
Le web invisible.....	42
2.3 <i>Le critère de fréquence</i> .....	42
3. ELABORATION D'UNE GRILLE DE SELECTION .....	44
3.1 <i>Les champs</i> .....	44
Identification et validation documentaire .....	44
Masque de saisie .....	45
Indications de suivi .....	46
3.2 <i>Proposition de grille de sélection</i> .....	47
<b>ARCHIVER LE WEB FOREZIEN.....</b>	<b>49</b>
1. EN AMONT, DEUX ORIENTATIONS POSSIBLES.....	49
1.1 <i>La tentation de l'exhaustivité</i> .....	50
Les collections .....	50
Les territoires du web.....	51
1.2 <i>La sélectivité : un choix plus raisonné ?</i> .....	52
Ces catégories sont-elles pertinentes ? .....	54
2. PRENDRE EN COMPTE LE DOCUMENT WEB DANS SA SPECIFICITE .....	55
2.1 <i>Dépasser l'aporie du contenu</i> .....	55
L'unité et l'écheveau .....	56
2.2 <i>Problématiser et constituer des corpus</i> .....	56
3. PREVOIR LA RESTITUTION .....	58
3.1 <i>Une offre structurée</i> .....	58
Quelle valorisation ? .....	59
Signalement .....	59
3.2 <i>La bibliothèque municipale prestataire ?</i> .....	60
Tisser des partenariats .....	60
Organiser l'accompagnement de l'utilisateur.....	61
<b>CONCLUSION.....</b>	<b>63</b>
<b>BIBLIOGRAPHIE .....</b>	<b>67</b>
ARCHIVER LE WEB : PROJETS ET METHODES .....	67
LE CADRE JURIDIQUE EN FRANCE .....	69
LA BNF ET LE DEPOT LEGAL DU WEB .....	70
POLITIQUE DOCUMENTAIRE ET FONDS REGIONAUX .....	71
<b>TABLE DES ANNEXES .....</b>	<b>75</b>

# Introduction

---

Deux obstacles majeurs ont longtemps freiné la collecte des objets numériques disponibles sur l'internet, pourtant appréhendée conjointement par la Bibliothèque nationale de France (BnF) et l'Institut national de l'audiovisuel (Ina) depuis 1998 : l'absence de cadre juridique et les questions techniques relatives à la conservation et à l'accessibilité des archives ainsi constituées. Depuis août 2006, la France s'est dotée d'une loi qui élargit le champ du dépôt légal à l'internet et attribue la mission de conservation à l'Ina pour les sites traitant du domaine audiovisuel, la BnF devant se charger de collecter tous les autres. Par ailleurs, si les problèmes liés à la durée de vie des supports numériques et à la rapide obsolescence des logiciels qui permettent d'accéder à l'information continuent de tenir les spécialistes de l'archivage en alerte, l'évolution des techniques a permis d'envisager sereinement la constitution d'une mémoire numérique. Quelques exemples en témoignent : certaines revues ont abandonné le tirage sur papier au profit de la seule forme électronique réalisant ainsi une baisse considérable des coûts et optimisant leur diffusion ; de son côté le département des Arts du spectacle de la BnF assure la capture régulière des sites web des théâtres de France, en portant une attention particulière aux newsletters et aux programmes sur fichier en *pdf*, en complément des *ephemera* sur support physique collectés depuis la fin du XIX<sup>ème</sup> siècle.

Néanmoins le débat autour d'une mémoire numérique ne fait que s'esquisser. Des initiatives publiques lui ont fait franchir le cercle des archivistes et des éditorialistes avertis en le portant sur la scène la plus large afin de lui restituer sa véritable dimension, celle d'un enjeu de société. Le très récent *Rize-centre Mémoires & Société* de Villeurbanne, qui associe dans une structure originale une médiathèque, les archives municipales et des espaces culturels et pédagogiques, a organisé au printemps 2008, conformément à sa vocation de « transmettre hier, [d'] imaginer demain », en partenariat avec les Archives municipales de Lyon, une exposition intitulée « Imaginons... une mémoire sans papier » dans laquelle trente-sept artistes contemporains ont livré sous forme d'oeuvres originales leur vision de la conservation de la mémoire à l'heure du numérique. Le production exponentielle de documents électroniques, dans tous les domaines de la société, transforme les missions traditionnelles des institutions chargées de la mémoire – stockage, conservation et valorisation des collections – en un défi permanent intrinsèquement lié à la rapide évolution des techniques et à la métamorphose des objets archivables.

Il nous semble que par le truchement de la création artistique, cette initiative permet d'éclairer la problématique de l'archivage du web à la fois dans ce qui la rapproche des savoir-faire traditionnels des bibliothèques et dans les aspects les plus inédits auxquels sont confrontés les professionnels. La simple migration des contenus d'un support physique vers une forme numérique ne pose guère question. Plus délicates à appréhender se révèlent les configurations d'informations nativement numériques telles que les hypertextes, les réseaux de sites, les bibliothèques électroniques, les forums, les blogs, etc. « *Cet espace virtuel – Web, Toile, cybermonde...*

– *n'est pas un simple reflet de l'espace réel, dont il dupliquerait les règles. C'est un continent spécifique (le sixième), avec ses logiques, son langage, ses impasses.*<sup>1</sup> » résumait récemment un journaliste. A la nécessité de prendre en compte cette architecture particulière, en forme de constellation, s'ajoute peut-être l'effort d'un décentrement volontaire pour qui veut approcher la réalité du web dans son rapport le plus immédiat avec le monde contemporain afin de ne pas « être conduit à prendre l'objet livre comme critère d'évaluation du texte électronique, plutôt que de se demander si celui-ci ne représenterait pas une forme plus perfectionnée de la textualité, désormais libérée d'un encapsulement qui ne faisait que répondre à des contraintes de transport et d'une linéarité imposée par la fixité de l'inscription.<sup>2</sup> » Patrick Bazin ajoute qu'il en va de même de la mémoire : « à trop l'assimiler à un stock – de bibliothèque ou de musée – auquel il s'agirait simplement d'accéder pour y prendre livraison de données intangibles, on risque de ne pas prendre toute la mesure du problème posé qui est celui de la transmission de la vie culturelle, ou plutôt de la transmission comme élément intrinsèque de la vie culturelle.<sup>3</sup> »

A cet égard, il nous semble que les bibliothèques municipales ont un rôle particulier à jouer dans le cadre d'une politique concertée de l'archivage du web dont la capture massive et la conservation sont désormais dévolues exclusivement à la BnF. Leur inscription dans les réseaux culturels régionaux et leur connaissance des acteurs locaux en font des partenaires précieux de la bibliothèque nationale qui intègre depuis plusieurs années à sa réflexion sur la collecte leur capacité d'expertise.

Cette étude voudrait proposer une approche méthodologique qui mette en évidence les enjeux de ce type de collecte et souligne les choix auxquels seront confrontés les bibliothécaires. En l'attente d'un décret fixant les conditions de la restitution des archives du web hors de la bibliothèque nationale, les scénarios imaginés sont de l'ordre de la projection. Cependant, les expérimentations menées par la BnF en collaboration avec les bibliothèques titulaires du dépôt légal imprimeur (BDLI) s'avèrent d'ores et déjà fécondes et prometteuses. Regarder l'archivage du web depuis l'échelon de la bibliothèque municipale ne revient pas seulement à opérer une reterritorialisation des objets numériques capturés en masse, qui ont par nature vocation à s'affranchir de la délimitation spatiale, c'est aussi éclairer une part invisible du web : l'écheveau constitué par la vie d'un territoire, son dynamisme, sa créativité, son réseau associatif, sa réactivité aux événements, la place de ses institutions, etc. Ce tableau est un choix conscient, à l'image de ce qu'est une collection<sup>4</sup> ; ne doutons pas qu'il devienne un jour l'objet d'une attente légitime des usagers.

Aussi serons-nous amenés à examiner l'intérêt d'un archivage de ce type et à esquisser un schéma des pratiques possibles en France, dans le respect du cadre juridique fixé par la loi de 2006, en nous fondant notamment sur le partenariat noué par la BnF avec les BDLI. Nous proposerons dans un second temps une méthodologie pour une politique d'archivage avant d'évaluer les options qui s'offrent aux bibliothécaires à travers l'observation d'un champ particulier, celui du web forézien, autour de la Médiathèque de Saint-Etienne.

---

<sup>1</sup> Jean-Claude GUILLEBAUD, « Une technologie définissante ? » in *Le Nouvel Observateur*, 23/08/2008.

<sup>2</sup> Patrick BAZIN, « La mémoire reconfigurée », in *Cahiers de médiologie*, n°11, 2001, p. 178.

<sup>3</sup> *Idem*.

<sup>4</sup> La collection se définit par un « regroupement volontaire de documents, d'objets, d'informations de provenances diverses, rassemblés en raison de la similitude d'un ou plusieurs de leurs caractères (contrairement au fonds d'archives qui est constitué organiquement) » rappelle le *Vocabulaire de la documentation*, coord. Arlette BOULOGNE, Paris : ADBS éd., 2004.

# Intérêt de l'archivage du web en bibliothèque municipale. Etat des pratiques.

---

## 1. POURQUOI ARCHIVER LE WEB REGIONAL ?

Une enquête menée auprès d'une soixantaine de bibliothèques municipales de France<sup>1</sup>, en préalable à cette étude, fait apparaître un intérêt variable pour l'archivage du web régional. Pourquoi mettre en place un système particulier de sélection des sites alors que la BnF se charge de la collecte globale et régulière du web français ? Comment la bibliothèque territoriale pourrait-elle faire face aux problèmes de stockage des documents numériques archivés et serait-elle en mesure d'en assurer la pérennité dans un contexte de développement exponentiel de l'internet ? A ces difficultés s'ajoutent le problème juridique, la formation technique du personnel et la mobilisation d'une équipe spécialisée dans la recherche et le suivi des sites. Autant d'obstacles qui apparentent ce projet à une entreprise herculéenne.

Néanmoins, la nécessité de proposer aux lecteurs des repères dans la consultation de l'internet, inséparable des fonctions de médiation et de préconisation du bibliothécaire, est rarement absente des réponses reçues. Nombre d'établissements ont proposé des signets sur leur portail ou bien un annuaire de sites permettant à l'utilisateur d'accéder à des sites documentaires ou institutionnels – de nature régionale ou nationale – dans une perspective d'extension de l'offre documentaire et de facilitation de l'accès aux services disponibles sur l'internet. Là encore, les déceptions sont perceptibles : force est de constater que ces ressources sont peu sollicitées et qu'elles ne correspondent plus guère aux usages de l'internaute, désormais expert en maniement de moteurs de recherche.

Si plusieurs de ces interrogations peuvent trouver une réponse satisfaisante dans une architecture nationale de l'archivage du web dont les initiatives de la BnF laissent imaginer la structure – nous y reviendrons –, il importe tout d'abord de se demander quels usages fondent la volonté de mettre en place un tel projet, et ce que la bibliothèque municipale doit entendre à travers l'expression de « web régional » afin de lever tout malentendu et de montrer en quoi cet objet relève de ses missions.

---

<sup>1</sup> Voir le questionnaire en Annexe 1.

## 1.1 Quelques statistiques sur la consultation de l'internet

### La consultation de l'internet en France

Rappeler que la consultation du web s'est rapidement imposée comme un moyen privilégié d'accès à l'information relève de l'évidence. Néanmoins les statistiques confirment l'augmentation impressionnante du nombre d'internautes, certainement sous l'effet de deux facteurs : la multiplication des connexions à haut débit et l'orientation « participative » du Web 2.0 qui permet de diversifier les formats de publication : wikis, blogs,... En juillet 2008, la société Médiamétrie évaluait à près de 32 millions le nombre d'internautes en France<sup>1</sup>, soit 60 % de la population, dont 5 % de plus qu'en juillet de l'année précédente. Les enquêtes internationales placent à la même époque la France au huitième rang mondial pour le nombre d'utilisateurs de l'internet par rapport à la population totale<sup>2</sup>. La même étude de Médiamétrie nous apprend que 14 millions de foyers possèdent un accès à l'internet et que 95 % des internautes à domicile disposent d'une connexion à haut débit.

### Le domaine français

Le domaine français de l'internet enregistre une croissance plus spectaculaire encore. L'Association Française pour le Nommage Internet en Coopération (AFNIC), qui gère la base de données des noms de domaine géographiques en *.fr* ou *.re* (La Réunion) évalue à près d'1,25 million le nombre de domaines délégués à la mi-octobre 2008. Au premier janvier de la même année, on en totalisait 991 000. Le nombre total de domaines a doublé depuis septembre 2006 où l'on en comptabilisait environ 618 000.<sup>3</sup> Ce tableau général qui met en valeur le dynamisme du web français mériterait d'être affiné par une étude régionale de la création des sites web. Or les statistiques demeurent très parcellaires en ce domaine puisque l'enregistrement ne prend pas en compte une donnée géographique de ce niveau qui ne serait du reste qu'indicative puisqu'un site consacré à une région peut fort bien être créé ailleurs. Et encore faudrait-il s'entendre sur la notion de « lieu de création » : s'agit-il du lieu de conception du site ou bien de la région qui héberge le serveur ?

---

<sup>1</sup> Pour Médiamétrie est définie comme internaute une personne âgée de 11 ans et plus et qui s'est connectée à l'internet au moins une fois au cours du mois, quel que soit le mode de connexion.

Source : [http://www.mediametrie.fr/resultats.php?rubrique=net&resultat\\_id=585](http://www.mediametrie.fr/resultats.php?rubrique=net&resultat_id=585) [Consulté le 13/10/2008]

<sup>2</sup> Source : <http://www.internetworldstats.com/top20.htm> [Consulté le 17/10/2008]

<sup>3</sup> Source : <http://www.afnic.fr/actu/stats/domaines/historique> [Consulté le 17/10/2008]

## Une étude des usages dans les régions françaises

Une étude des usages de l'internet en région peut s'avérer plus significative. On observe dans un premier temps un phénomène attendu : ce sont les régions françaises les plus peuplées qui génèrent le plus grand nombre de visites. L'Ile-de-France, région qui concentre environ deux fois plus d'habitants que Rhône-Alpes, enregistre deux fois plus de visites. Dans le tableau ci-dessous, l'indicateur de visites prend pour indice 1 le nombre de visites enregistrées depuis le Limousin (région où elles sont les moins nombreuses)<sup>1</sup>. Ainsi, durant la semaine d'observation, pour 1 visite effectuée depuis cette région, 1801 sont réalisées depuis l'Ile-de-France.

Région	Visites	Habitants
1 Ile-de-France	1801	11 491 000
2 Rhône-Alpes	988	6 005 000
3 Provence Alpes Côte d'Azur - Corse	788	5 060 000
4 Pays de la Loire	494	3 426 000
5 Nord-Pas-de-Calais	478	4 043 000
6 Aquitaine	467	3 099 000
7 Bretagne	428	3 081 000
8 Alsace	385	1 817 000
9 Midi-Pyrénées	372	2 755 000
10 Languedoc-Roussillon	369	2 520 000
11 Poitou-Charentes	288	1 713 000
12 Bourgogne	286	1 624 000
13 Lorraine	272	2 339 000
14 Basse-Normandie	217	1 449 000
15 Picardie	166	1 886 000
16 Champagne-Ardenne	165	1 339 000
17 Haute-Normandie	164	1 811 000
18 Auvergne	161	1 334 000
19 Centre	61	2 505 000
20 Franche-Comté	42	1 146 000
21 Martinique	35	399 000
22 Guadeloupe	35	447 000
23 Guyane	14	202 000
24 La Réunion	7	784 000
25 Limousin	1	725 000

### Nombre de visites hebdomadaires par région par rapport au nombre d'habitants (nombre de visites depuis le Limousin : indice 1)

<sup>1</sup> Etude réalisée par XiTiMonitor du 23/04/2007 au 29/04/2007.

Source : <http://www.xitimonitor.com/pdf/fr-FR/Régions%20françaises%20-%20Mai%202007.pdf> [Consulté le 17/10/2008]

Région	Visites	Habitants
1 Alsace	154	1 817 000
2 Bourgogne	128	1 624 000
3 Poitou-Charentes	122	1 713 000
4 Rhône-Alpes	119	6 005 000
5 Ile-de-France	114	11 491 000
6 Provence Alpes Côte d'Azur - Corse	113	5 060 000
7 Aquitaine	109	3 099 000
8 Basse-Normandie	108	1 449 000
9 Languedoc-Roussillon	106	2 520 000
10 Pays de la Loire	105	3 426 000
11 Bretagne	101	3 081 000
12 Midi-Pyrénées	98	2 755 000
13 Champagne-Ardenne	89	1 339 000
14 Auvergne	87	1 334 000
15 Nord-Pas-de-Calais	86	4 043 000
16 Lorraine	84	2 339 000
17 Haute-Normandie	66	1 811 000
18 Picardie	64	1 886 000
19 Martinique	63	399 000
20 Guadeloupe	56	447 000
21 Guyane	50	202 000
22 Franche-Comté	27	1 146 000
23 Centre	18	2 505 000
24 La Réunion	6	784 000
25 Limousin	1	725 000

**Nombre de visites hebdomadaires par habitant et par région  
(nombre de visites depuis le Limousin : indice 1)**

Si le nombre de visites global par région reste lié à la densité de la population, le relevé de la fréquentation de l'internet par habitant révèle un classement plus inattendu. Les trois régions classées en tête (Alsace, Bourgogne, Poitou-Charentes) sont respectivement treizième, seizième et quinzième par nombre d'habitants.<sup>1</sup> Suivent les trois régions les plus peuplées déjà repérées comme les plus demandeuses de l'internet. On peut remarquer enfin que les régions Pays-de-la-Loire et Nord-Pas-de-Calais, régions très peuplées arrivant au quatrième et au cinquième rang de la demande globale, présentent un taux de fréquentation individuelle plus modeste (105 sites par habitant pour la première, 86 pour la seconde).

A défaut de pouvoir évaluer quantitativement la production du web relative à chaque région, nous pouvons constater que les indices de fréquentation montrent l'importance de ce média dans la sphère régionale. Les chiffres relatifs à la consultation individuelle soulignent la demande qui émane du territoire tout entier, sans que la densité de la population ou la répartition des concentrations urbaines soient des facteurs déterminants. La place prépondérante de l'internet dans l'accès à l'information et le positionnement de l'internaute comme producteur de documents variés suffisent à légitimer la démarche d'archivage du web, et notamment pour une bibliothèque municipale, la conservation de la part du web d'intérêt local.

<sup>1</sup> Source INSEE au 01/01/2007.

## 1.2 Qu'entend-on par « web régional » ?

Une étude menée sur les conditions d'archivage et de conservation des sites web en bibliothèque municipale propose une triple acception pour l'expression « web régional ». Elle pourrait désigner les sites web créés dans la région, ceux hébergés par un serveur local ou bien encore les sites ayant un intérêt documentaire lié au territoire.<sup>1</sup> L'auteur arrête son choix à la formule « web d'intérêt régional » qui met l'accent sur le contenu informationnel ou patrimonial à caractère régional de l'objet numérique. Il va de soi que nous faisons nôtre cette définition que nous tenterons de préciser avant de proposer plus loin l'exemple d'une sélection de sites dans le cadre d'un projet d'archivage du web forézien.<sup>2</sup>

### Essai de définition : l'objet et la méthode

Nous pouvons considérer qu'appartient au « web régional » tout objet numérique, quel qu'en soit le format, dont le contenu renvoie explicitement et majoritairement à une ou plusieurs données régionales de nature statistique, institutionnelle, historique, géographique, artistique, économique, politique ou sociologique. Nous incluons aussi bien les documents dont la volonté descriptive du créateur est explicite (les sites « portant sur » un élément régional, comme les sites d'histoire locale) que les objets pouvant être considérés par l'utilisateur du web comme le reflet d'une dimension régionale marquée (ainsi le site d'une entreprise dont la notoriété assure la représentation d'un lieu, d'une population). Il nous semble donc qu'une première qualité du « sélectionneur<sup>3</sup> » est de chercher à adopter le regard du « créateur » de l'objet – identifiable ou non – afin de découvrir une intention, et de se projeter dans la perspective du chercheur qui étudie la vie régionale, avec un œil critique.

A ce stade, deux observations s'imposent. Les champs couverts par les sites d'intérêt historique, géographique, artistique, économique, politique ou sociologique se situent dans le prolongement des principaux secteurs d'étude déjà représentés dans les fonds locaux des bibliothèques municipales. La continuité des collections, à travers les nouveaux supports numériques, est un élément essentiel à la recherche des sites web d'intérêt régional. Par ailleurs, si le critère d'« intentionnalité » reste opératoire dans le choix d'un grand nombre d'objets, il n'en va pas de même du « regard critique » censé mimer la démarche sélective du chercheur. Il serait hasardeux de parier sur les centres d'intérêt qui préoccuperont les générations à venir et inconsideré de s'appuyer sur de simples projections en définitive tributaires de critères contemporains. Aussi proposons-nous d'entendre par « regard critique » la sélection de sites dont la qualité documentaire et la surface informationnelle offrent un matériau fiable et exploitable à l'usager des bibliothèques.

---

<sup>1</sup> Magali HAETTIGER, *L'Archivage des sites web d'intérêt régional*, Mémoire d'étude, Villeurbanne : ENSSIB, 2003, p. 51.

<sup>2</sup> Voir l'annexe 3.

<sup>3</sup> C'est le terme employé par la BnF lors des campagnes de collecte des sites électoraux.

## Une démarche consciente de constitution d'une collection

La démarche que nous proposons est fondée sur le constat que le « web régional » est une entité qui n'a pas d'existence propre, à la différence du domaine en *.fr* par exemple. Il s'agit de la délimitation intellectuelle d'un champ de l'internet suivant des critères de sélection, dont la qualité repose sur deux facteurs essentiels : la capacité d'expertise des bibliothécaires et leur aptitude à assurer une veille documentaire. A l'instar d'une collection, le « web régional » constitué par une bibliothèque regroupe des objets divers aux formats variés dont le seul point commun est de présenter un intérêt local.

Entendue de cette manière, la collection numérique présente une unicité de support mais s'avère destinée à intégrer tous les champs couverts par les fonds régionaux. Les exemples abondent. Ainsi, de nombreuses municipalités au budget modeste pallient l'absence de publication communale par un site web parfois entretenu grâce à un bénévole. La liste des 327 communes de la Loire présentée sur le site du Conseil général<sup>1</sup> permet de recenser 85 sites web municipaux dont très peu se doublent d'une publication papier. Un recensement déjà ancien de la Chambre de Commerce et d'Industrie de Saint-Etienne évaluait à 40,9 % en 2003 la proportion des entreprises du département disposant d'un site web. Les pages des principales institutions culturelles proposent parfois une courte vidéo des spectacles qu'elles accueillent.

## Un objet d'étude autonome

Enfin, il va de soi que la documentation numérique regroupée peut constituer un objet d'étude autonome. Témoin de la vitalité économique et culturelle d'une région, il transmet parfois l'image de la réactivité d'une population à l'occasion d'un fait divers marquant ou bien d'un temps fort de la vie citoyenne. Les formes participatives du web (blogs, wikis, etc.) peuvent faire l'objet d'une sélection dans cette perspective. Une bibliothécaire de la Réunion<sup>2</sup> résume l'intérêt des archives du web régional : elles rendent compte du dynamisme du web local, participent à la connaissance du territoire et de ses acteurs et peuvent être rendues aisément accessibles au lecteur grâce à la tradition d'indexation et de classement des bibliothèques. Remarquons enfin qu'il n'y a pas coïncidence entre l'espace régional et les « territoires du web » qui adoptent une structuration spécifique et recouvrent des formes d'expression multiples : ces caractères fournissent des sources d'études à de nombreux chercheurs en sciences sociales et en sciences de l'information et de la communication.<sup>3</sup>

---

<sup>1</sup> [http://www.loire.fr/display.jsp?id=j\\_4149](http://www.loire.fr/display.jsp?id=j_4149) [Consulté le 18/11/2008]

<sup>2</sup> Courrier reçu dans le cadre de l'enquête mentionnée à l'Annexe 1.

<sup>3</sup> Nous pouvons, par exemple, renvoyer à l'équipe « Réseaux, Savoirs & Territoires » de l'Ecole Normale Supérieure, coordonnée par Eric GUICHARD. Le programme de recherche est disponible en fichier pdf : <http://barthes.ens.fr/equipe/QuadriERST2005.pdf> [Consulté le 18/11/2008].

## 1.3 Intérêt du web régional en bibliothèque municipale

### Enrichissement et continuité des collections

En situant la constitution d'une collection d'archives numériques dans la continuité d'une politique documentaire, nous considérons qu'elle s'inscrit dans la mission de collecte et d'actualisation des fonds régionaux dévolue aux grandes bibliothèques municipales, et notamment aux bibliothèques municipales à vocation régionale (BMVR).<sup>1</sup> La migration de l'information du support imprimé traditionnel vers le numérique, dans un souci de réduction des coûts et dans une optique de plus large diffusion, contraint l'acquéreur à se tourner vers le nouveau support. Le cas est fréquent pour les revues savantes, les organes de diffusion syndicale (le militant est parfois invité à imprimer le fichier numérique, voire à transformer le tract virtuel en affiche, au besoin), les bulletins associatifs et municipaux des petites communes, etc. Ainsi, le transfert d'une publication régulière vers l'internet métamorphose *de facto* le périodique en « éphémère » et pose le problème de la continuité de la collection en bibliothèque et de sa conservation. Le risque est grand de voir disparaître avec le site hébergeant les fichiers toutes les archives du périodique numérique. Une capture régulière du site web, intégrant une périodicité et une profondeur adaptées, constitue une garantie contre ce risque de plus en plus souvent constaté en bibliothèque municipale, que l'absence fréquente de veille qualifiée sur l'internet ne fait qu'accroître. Plus volatiles encore semblent les formats de consultation tels que les blogs soumis à un hébergeur : que penserait-on par exemple de la perte d'un blog militant inscrivant une expérience individuelle dans un débat de société ?<sup>2</sup>

### Valorisation de la bibliothèque

Non seulement l'archivage du web régional constitue le prolongement naturel de l'acquisition de la documentation locale, mais il est aussi un moyen de valorisation de la bibliothèque auprès de plusieurs catégories de public, du simple curieux au chercheur. En cela non plus il ne fait pas figure de nouveauté radicale, puisque de nombreux établissements mettent déjà à la disposition de leurs usagers des liens vers un choix de sites « vivants » significatifs. La bibliothèque Méjanes-Cité du Livre d'Aix-en-Provence opère une sélection et procède à un classement de sites régionaux (non commerciaux et non personnels) retenus pour leur fiabilité et la qualité de l'information.<sup>3</sup> Un critère déterminant est leur lien avec les fonds spécifiques dont dispose la bibliothèque. Ainsi des signets associent des sites de l'internet à la bibliothèque d'art, au fonds provençal, au pôle Méditerranée, aux fonds Cézanne, Saint-John-Perse, Camus. D'autres renvois dirigent le lecteur vers des pages liées à la vie politique, administrative et sociale de la région.

---

<sup>1</sup> Nous renvoyons à l'étude de Jean-Christophe BROCHARD, *Après les inaugurations, quel avenir pour les missions régionales des BMVR ?* Mémoire de DCB, ENSSIB, Villeurbanne, 2005.

<sup>2</sup> Songeons par exemple, en dehors de toute référence strictement régionale, au blog militant de Jean-Paul Brighelli, conçu comme un commentaire quasi immédiat de l'actualité liée aux problématiques de l'éducation. Ce blog est le prolongement « au jour le jour » d'ouvrages de son créateur ayant connu des succès de librairie.  
<http://bonnetdane.midiblogs.com> [Consulté le 18/10/2008]

<sup>3</sup> Courrier reçu dans le cadre de l'enquête mentionnée à l'Annexe 1.

La bibliothèque municipale de Lyon (BML), partenaire de la BnF durant les campagnes de collecte du web électoral, a tiré profit de son travail de recherche et d'analyse des sites locaux et créé un « dossier repère »<sup>1</sup> sur les scrutins municipaux en Rhône-Alpes intégré à la rubrique « Points d'Actu » de son portail. Un ciblage ponctuel de l'internet, aussi exhaustif que possible, à l'occasion de l'élaboration de dossiers d'actualité, vient nourrir ce service aux lecteurs offert par la BML.<sup>2</sup> La bibliothèque municipale de Lille, associée elle aussi aux collectes de la BnF, a intégré à son catalogue les liens permettant d'accéder aux sites référencés par son équipe.

On peut ainsi considérer que l'accès aux archives du web viendrait à l'appui des services thématiques aux lecteurs imaginés par certaines bibliothèques et serait de façon plus générale en mesure de valoriser les collections dont les établissements sont les dépositaires. Un usager averti aura la possibilité de consulter les versions précédentes d'un site enregistré par Internet Archive grâce à la *Wayback Machine*<sup>3</sup> mais pourra-t-il se satisfaire de ces captures aléatoires de l'objet numérique alors que les autres ressources dont il dispose dans l'environnement de la bibliothèque bénéficient de la capacité d'expertise documentaire des professionnels ?

## 2. LE CADRE JURIDIQUE ACTUEL

Avant d'examiner les schémas possibles de collecte et de valorisation du web régional, il importe de préciser le cadre légal dans lequel s'effectue actuellement l'archivage de l'internet en France.<sup>4</sup> Seule la mission de dépôt légal dévolue à la BnF permet un archivage par capture, le volume considérable d'objets numériques à conserver rendant totalement inadapté le processus traditionnel de dépôt par l'imprimeur et l'éditeur, ou bien par le « créateur » d'un site, si l'on cherche à transposer cette procédure au web.

### 2.1 Le dépôt légal des sites

La loi intitulée « Droit d'auteur et droits voisins dans la société de l'information » (dite DADVSI), parue au Journal Officiel du 3 août 2006, est issue de la transposition d'une directive européenne de 2001.<sup>5</sup> Le titre IV traite du dépôt légal et comprend les articles 39 à 47 qui modifient le Code du patrimoine et d'autres textes de loi. Les dispositions suivantes sont indiquées.

Le dépôt légal est étendu aux logiciels et aux bases de données. Sont également soumis à cette obligation « les signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique » (art. 39). La formule employée est empruntée au Code des postes et des communications téléphoniques modifié en 2004 et elle désigne explicitement l'internet.

L'article 41 précise les modalités susceptibles d'être employées par les organismes dépositaires pour assurer leur mission de dépôt légal : « *Ils peuvent procéder eux-mêmes*

---

<sup>1</sup> [http://www.pointsdactu.org/article.php?id\\_article=863](http://www.pointsdactu.org/article.php?id_article=863) [Consulté le 18/11/2008]

<sup>2</sup> Courrier reçu dans le cadre de l'enquête mentionnée à l'Annexe 1

<sup>3</sup> <http://www.archive.org/web/web.php> [Consulté le 18/10/2008]

<sup>4</sup> Cf. ILLIEN, Gildas ; GAME, Valérie, « Le dépôt légal d'Internet à la Bibliothèque nationale de France : Cadre juridique, modèle de collecte, évolutions des métiers », *BBF*, 2006, n° 3, p. 82-85. [en ligne] <http://bbf.enssib.fr> [Consulté le 02/11/08]

<sup>5</sup> Directive européenne 2001/29/CE.

à cette collecte selon des procédures automatiques ou en déterminer les modalités en accord avec ces personnes. La mise en oeuvre d'un code ou d'une restriction d'accès par ces personnes ne peut faire obstacle à la collecte par les organismes dépositaires précités. » Le législateur a eu le souci d'adapter les conditions du dépôt légal à la complexité de l'internet qui rend inadéquate la démarche traditionnelle de communication des documents par l'éditeur.<sup>1</sup> La collecte par aspiration est donc autorisée et les organismes chargés de la gestion des noms de domaine peuvent communiquer aux organismes dépositaires les données d'identification dont ils disposent. Ainsi la BnF reçoit régulièrement de l'AFNIC la liste des enregistrements des domaines en *.fr* et *.re*. Les organismes chargés du dépôt légal de l'internet sont l'Institut national de l'Audiovisuel (INA) et la BnF avec la coopération du conseil supérieur de l'audiovisuel et les gestionnaires de noms de domaine.

## 2.2 Propriété intellectuelle et droit moral : la consultation des archives

En contrepartie de l'exception au droit d'auteur introduite par la loi afin de permettre l'aspiration des sites sans requérir l'accord des ayants droit, l'article 42 restreint les conditions de consultation des archives de l'internet aux conditions suivantes : «*sur place par des chercheurs dûment accrédités par chaque organisme dépositaire sur des postes individuels de consultation dont l'usage est exclusivement réservé à ces chercheurs* ». Le même article modifie le code du patrimoine en autorisant «*la reproduction d'une oeuvre, sur tout support et par tout procédé, lorsque cette reproduction est nécessaire à la collecte, à la conservation et à la consultation sur place* » dans les conditions définies. Le législateur a cherché à tenir compte des conditions spécifiques de la conservation et de la consultation d'archives numériques dépendantes d'un environnement informatique en évolution : en effet, le transfert d'un document vers un autre support pouvait être considéré comme une atteinte au droit moral de l'auteur en ce qui concerne les supports traditionnels.

Pour le cas des documents sonores télévisés ou radiodiffusés, dont la collecte peut présenter un intérêt pour les bibliothèques (que l'on songe aux messages vidéos délivrés par les candidats lors des dernières campagnes électorales, ou bien encore aux vidéos disponibles sur *YouTube* ou *DailyMotion* qui provoquent parfois des mouvements d'opinion), l'INA est «*seul responsable de la collecte* » ; de plus, «*il participe avec la Bibliothèque nationale de France à la collecte, au titre du dépôt légal, des signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication publique en ligne*» (article 45).

---

<sup>1</sup> Gildas ILLIEN note que la législation britannique «*oblige nos homologues d'outre-Manche à obtenir l'accord des ayants droit d'un site avant sa capture.* » Ainsi, «*la British Library estime que, sans modification de la loi, le dispositif actuel ne lui permettra pas de couvrir plus de 0,6 % du domaine britannique de l'internet d'ici dix ans.* » In «*Le dépôt légal de l'internet en pratique : les moissonneurs du web* », *BBF* n°6, 2008, p. 21.

## 2.3 Quel décret d'application ?

L'article 41 annonce un décret d'application en Conseil d'Etat permettant de fixer «*les conditions de sélection et de consultation des informations collectées [...] après avis de la Commission nationale de l'informatique et des libertés.*» Ce décret peut être d'une importance capitale dans la mise en place d'une structure d'archivage concertée du web entre l'organisme chargé du dépôt légal, seul habilité à effectuer les aspirations du web, et les bibliothèques municipales prescriptrices.<sup>1</sup> On conçoit mal comment ces dernières pourraient consacrer des moyens à l'accomplissement de ces tâches sans l'espoir d'offrir à leurs usagers l'accès à ces archives et sans la possibilité de valoriser ce travail accompli en matière de constitution d'une documentation régionale. Aussi, forte de son expérience de collaboration avec les BDLI à l'occasion de plusieurs campagnes ciblées de collecte du web électoral, la BnF s'est engagée à proposer l'extension du droit de consultation des archives aux bibliothèques titulaires du dépôt légal imprimeur en région. De cette manière, à défaut de pouvoir trouver à court terme un point d'accès en bibliothèque municipale, il serait possible de consulter dans chaque région les archives de l'internet.<sup>2</sup>

## 3. LA BNF EN CAMPAGNE : RETOUR SUR UNE EXPERIENCE DE COOPERATION AVEC LES BIBLIOTHEQUES

### 3.1 Une démarche patrimoniale et expérimentale

#### De la collecte large à la collecte thématique

La BnF s'est préparée à assumer l'extension de sa mission pluriséculaire de dépôt légal aux objets numériques de l'Internet bien avant la loi du 1<sup>er</sup> août 2006. Depuis 1999, des expérimentations ont été menées dans la perspective d'un archivage automatique de grande ampleur. Il s'est agi tout d'abord d'évaluer la capacité des robots à collecter les unités complexes que sont les sites : de nombreux obstacles techniques peuvent entraver la progression vers les pages, les fichiers encapsulés et les liens qu'ils contiennent. Dans le cadre du projet européen NEDLIB<sup>3</sup>, une première collecte automatique a été testée, puis en 2002 le domaine en *.fr* a fait l'objet d'une capture.

---

<sup>1</sup> Lionel MAUREL fait remarquer que le rapport du conseiller François Stasse, préalable à la réflexion sur la loi, contenait des « propositions intéressantes » qui « n'ont malheureusement pas été retenues. » Il envisageait notamment d'étendre l'exception modifiant le Code du patrimoine à « toute bibliothèque publique atteignant un niveau de rayonnement à préciser d'un commun accord. » Il faisait également remarquer que la directive européenne de 2001 n'imposait pas de limiter la consultation sur place aux « chercheurs dûment accrédités » et il ménageait la possibilité d'une mise en réseau des bibliothèques qui « auraient pu proposer aux usagers la consultation de leur catalogue de façon à ce qu'à partir d'une bibliothèque, il soit possible d'accéder au fonds éditorial de n'importe quelle autre bibliothèque » (F. Stasse). In *Bibliothèques numériques : le défi du droit d'auteur*, Presses de l'ENSSIB : Villeurbanne, 2008, p. 109.

<sup>2</sup> A cette date, un projet de décret a été déposé auprès du Conseil d'Etat.

<sup>3</sup> *Networked European Deposit Library* : projet réalisé par la Commission européenne entre 1998 et 2001 afin de créer un ensemble d'outils facilitant la mise en œuvre de dépôts numériques. Le logiciel « moissonneur » mis au point pour l'archivage des données a été remplacé par le robot Heritrix développé par le consortium IIPC (*International Internet Preservation Consortium*).

Rapidement le constat s'est imposé que la collecte permettait de prendre un cliché instantané du domaine français, sans toutefois constituer une collection patrimoniale satisfaisante. A l'évidence, la collecte automatique ne touchait que la « surface » de la Toile. Aussi, la BnF s'est attachée à l'élaboration d'outils spécialisés permettant de définir une cartographie des domaines à archiver (domaines nationaux et génériques en .org, .edu, .com, etc.), tout en repérant les fréquences de mise à jour ainsi que la durée de vie moyenne des sites visés. Il devenait aussi nécessaire de créer le signalement de données inaccessibles éventuelles qui feraient l'objet d'une collecte spécifique. Une telle ambition n'avait de sens qu'à l'échelle internationale, qui est aussi celle du web : c'est pourquoi le Consortium IIPC<sup>1</sup> (*International Internet Preservation Consortium*) développe la coopération d'institutions de recherche et d'établissements patrimoniaux autour de l'élaboration d'outils de capture, d'indexation et d'archivage du web en menant en parallèle une réflexion sur le traitement documentaire des objets numériques collectés et conservés. Créé en 2003 par onze bibliothèques nationales, le consortium est piloté et coordonné par la BnF. Cette collaboration scientifique et technique a permis la mise au point du robot de capture Heritrix utilisé par la BnF à l'occasion des campagnes que nous allons aborder.<sup>2</sup>

Malgré ces progrès dans la manière d'appréhender le dépôt légal du web, il s'est avéré qu'une part significative de la Toile, à valeur patrimoniale, échappait à ces captures à grande échelle. La création de sites événementiels, qui surgissent à l'occasion de rendez-vous politiques importants ou de faits d'actualité marquants<sup>3</sup> a vocation à entrer dans les archives du web ; elle est aussi éminemment volatile. Un autre type de prospection documentaire restait à envisager : celui d'une collecte thématique, ciblée, indiquant au cas par cas au robot la fréquence et la profondeur de la capture pour chaque site. Dans cette perspective, la démarche individuelle d'expertise et de préconisation requiert l'intervention du bibliothécaire : il s'agit d'une capture « manuelle ».

## **La campagne électorale, objet patrimonial et champ d'expérimentation**

La BnF a mis à profit l'opportunité de campagnes électorales à deux reprises avant l'année 2007. Les élections présidentielles et législatives de 2002 ont été l'occasion de saisir 1900 sites en 6300 captures, pour un volume archivé de 530 Go et 12 millions de fichiers. La « surprise » du premier tour de l'élection présidentielle, le 21 avril 2002, a laissé percevoir l'émergence d'une forme de réactivité des internautes par le biais de « forums » de discussion qui ont fait l'objet d'une collecte à leur tour. En 2004, les élections régionales et européennes ont donné lieu à une saisie plus fine, prenant en compte l'expérience précédente : les captures ont été de meilleure qualité et les volumes archivés moins importants (161 Go et 6,3 millions de fichiers pour les régionales). Le traitement documentaire des sites capturés a été notablement amélioré dans la perspective d'un archivage à long terme.

---

<sup>1</sup> Cf. LUPOVICI, Christian. *The international Internet Preservation Consortium : General strategy and state of work* [en ligne]. In Conference of European National Librarians, Helsinki., 27 septembre 2007. Disponible sur : <http://netpreserve.org/events/IIPC%20CENL%20070927.ppt> [Consulté le 18/10/2008]

<sup>2</sup> InternetArchive. *Heritrix*. Disponible sur : <http://crawler.archive.org/> [Consulté le 18/10/2008]

<sup>3</sup> La Bibliothèque du Congrès (Washington) a répertorié 30 000 sites, de tous pays, en lien avec les attentats du 11 septembre 2001.

C'est dans la continuité de cette expérimentation que la BnF a lancé en 2007 deux projets. Le premier a été conçu comme un ultime test avant la systématisation de collectes de grande ampleur, et le second comme une tentative de restitution auprès d'un public de chercheurs bien délimité. Le projet « Internet en campagne » s'inscrit tout naturellement dans la consitution de « *sources nouvelles pour l'histoire politique et sociale de la nation ainsi que pour la sociologie électorale* »<sup>1</sup> amorcée lors des campagnes précédentes. Son intérêt réside aussi dans la mise en place d'une « infrastructure permanente de collecte<sup>2</sup> » associant étroitement les bibliothèques titulaires du dépôt légal imprimeur (BDLI) à la bibliothèque nationale. Comme le souligne un rapport interne à la BnF<sup>3</sup>, l'étalement du projet d'octobre 2006 à juillet 2007 a permis d'éprouver la capacité tant logistique que technique du dispositif face à une « montée en charge » : le nombre de sites web a augmenté graduellement au cours des mois de campagne et la BnF a dû faire face à un accroissement proportionnel de ses charges tant dans le domaine de la collecte que dans celui de l'indexation par URL ou bien encore afin d'assurer l'archivage à long terme. Plus encore qu'en 2004, les BDLI ont pris une part active à ce projet et la coopération mise en place semble prometteuse dans la double perspective d'une sélection plus fine des sites à archiver, et d'un archivage concerté au plan national.

## 3.2 Méthodologie : une entreprise associant les BDLI

### La prise en compte de l'échelle locale

L'approche méthodologique d'une « web-campagne » électorale se devait de prendre en compte l'échelle locale de l'internet puisque c'est généralement dans cette sphère que les initiatives citoyennes sont les plus nombreuses et la vivacité des échanges entre militants constitue un sismographe de première importance sur le déroulement d'un processus électoral. A cette justification d'ordre politique ou sociologique, peut-être plus pertinente encore pour les élections législatives qui s'annonçaient après les présidentielles, s'ajoutait la volonté pour la BnF de relancer la coopération « extrêmement précieuse »<sup>4</sup> de « sélectionneurs sur le terrain »<sup>5</sup> déjà très appréciée lors des élections régionales et européennes de 2004. Leur ancrage dans le contexte régional et leur connaissance des acteurs locaux en faisaient des partenaires indispensables dans un projet de couverture du territoire le plus complet possible. Encore fallait-il auparavant élaborer une méthodologie pour la sélection des sites adaptée à un traitement manuel par des agents non spécialisés dans la collecte de l'internet et qui réponde aux critères techniques lisibles par le robot moissonneur. Huit bibliothèques titulaires du dépôt légal imprimeur en région se déclarèrent volontaires : Caen, Dijon, Lille, Limoges, Lyon, Nouméa, Poitiers et Strasbourg (Bibliothèque nationale universitaire).

---

<sup>1</sup> « Dossier de presse », Bnf. Disponible en ligne : [http://www.bnf.fr/pages/presse/dossiers/internet\\_campagne.pdf](http://www.bnf.fr/pages/presse/dossiers/internet_campagne.pdf) [Consulté le 26/10/2008]

<sup>2</sup> Clément OURY. *Elections 2007 : Bilan de l'organisation du projet*. Paris : Bibliothèque nationale de France, Direction des Services et des réseaux, département de la bibliothèque numérique, 28 septembre 2007, p. 5.

<sup>3</sup> *Idem*, p. 5.

<sup>4</sup> *Ibidem*, p. 27.

<sup>5</sup> *Ibidem*, p. 27.

## Les BDLI : partenariat et accompagnement

Outre les réunions organisées à la BnF, deux types d'outils furent créés à l'intention des sélectionneurs en région : des outils d'accompagnement et de discussion sous la forme d'une liste de diffusion (gérée par la BNU de Strasbourg), de comptes rendus systématiques des comités de sélection, de contacts téléphoniques ou par courriel avec les coordinateurs, et un outil de saisie en ligne des sites retenus par les agents en région. Il est intéressant de noter que les sélectionneurs des BDLI étaient en même temps affectés aux services les plus divers de leur bibliothèque : patrimoine, acquisitions/entrées, service informatique, avec une très légère majorité pour les activités les plus attendues dans ce domaine : dépôt légal imprimeur et fonds local ou régional.<sup>1</sup> On peut en déduire que si une sensibilisation au fonctionnement de l'internet et une pratique régulière sont indispensables, les activités de sélection ne nécessitent pas un usage expert de l'outil web et qu'elles ne s'éloignent guère des activités bibliothéconomiques traditionnelles liées à l'acquisition de la documentation. Des formations liées à la compréhension de l'architecture d'une adresse URL et d'un site web ont été dispensées à une vingtaine de sélectionneurs. Les indications principales sont reprises dans le « manuel de l'utilisateur » de la plateforme web de saisie des propositions de collecte,<sup>2</sup> afin de rendre compréhensibles les notions de *fréquence* et de *profondeur* nécessaires au robot collecteur.

### La tâche des sélectionneurs

L'instrument de saisie mis à la disposition des BDLI résume à lui seul les compétences qui entrent en jeu dans la préconisation des sites. Après avoir renseigné le *type d'élection* et exposé éventuellement, de manière brève, les raisons qui ont amené à retenir ce site, le sélectionneur était invité à valider le *type de site* correspondant à une typologie documentaire précise en sept catégories arrêtée conjointement par la BnF et ses partenaires autour de deux subdivisions : « les candidats et leur organisation » et « regards et opinions sur la campagne », en somme les sites *de* et *sur* la campagne. Venaient ensuite le nom du candidat et celui de sa formation contenu dans un menu déroulant, puis le choix éventuel de l'abonnement à la lettre d'information du site. Restaient enfin les critères plus délicats de *fréquence* et *profondeur* donnant lieu aux options suivantes :

<i>fréquence</i>	1 – une fois par semaine 2 – une fois par mois 3 – une fois avant, entre et après les scrutins 4 – une seule fois
------------------	--

<i>profondeur</i>	1 – domaine 2 – hôte ( <i>host</i> ) 3 – page 4 – page et contexte 5 – chemin ( <i>path</i> )
-------------------	---

<sup>1</sup> *Ibidem*, p. 27.

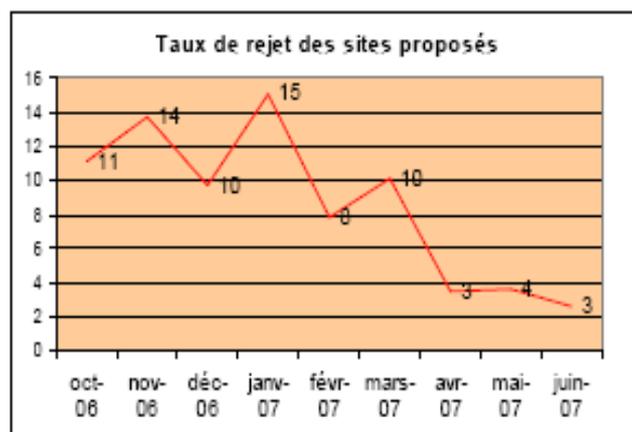
<sup>2</sup> Catherine LAGOUTE, Clément OURY. *Elections 2007 : réunion des sélectionneurs des BDLI (30/03/2007) – manuel utilisateur de l'outil*. Paris : Bibliothèque nationale de France, Direction des Services et des réseaux, département de la bibliothèque numérique, pp. 7-9.

En même temps qu'il donnait des indications sur le choix de la fréquence et de la profondeur en fonction des types de sites, le manuel de l'utilisateur proposait quelques exemples et recommandations notamment au sujet des « cas particuliers » que sont les sites de presse (doit-on saisir les articles à l'unité ?) et les blogs (quelle fréquence et quelle profondeur ?). La capture d'une vidéo de *Dailymotion* était expliquée.

Une fois les préconisations envoyées, il revenait à des agents du Département de la Bibliothèque numérique (DBN) d'opérer la validation technique des URL. Il s'agit à ce stade de vérifier que l'adresse URL est la plus pertinente afin d'éviter les doublons et de manière à écarter les adresses qui présenteraient des difficultés techniques pour le robot. La planification et la collecte sont les étapes suivantes d'un circuit qui doit permettre de capturer les sites les plus adaptés aux critères élaborés avant l'indexation et la sauvegarde. Le schéma directeur de ce processus est donné en annexe.<sup>1</sup>

## Éléments de bilan

Si les correspondants en BDLI se sont déclarés satisfaits des modes d'accompagnement proposés, la BnF peut-elle aussi tirer un bilan très positif de la collaboration avec ses partenaires. À cet égard, la courbe<sup>2</sup> représentant le taux de rejet des sites proposés entre octobre 2006 et juin 2007 est révélatrice de l'acquisition des compétences par les agents chargés du repérage des sites en région :



Le rapport de la BnF qui dresse le bilan organisationnel du projet cite un bibliothécaire de BDLI : ce modèle d'organisation, qui a permis la réalisation d'un cycle entier d'archivage du web, de la sélection à la restitution, lui semble de nature à permettre « l'enrichissement professionnel sur des missions courtes, précises et complémentaires des activités traditionnelles. »<sup>3</sup> C'est cette complémentarité qui nous semble l'acquis le plus précieux de l'expérimentation dans la perspective d'une collaboration pérenne entre la BnF et les bibliothèques municipales afin de constituer un réseau coordonné d'archivage du web régional.

<sup>1</sup> Voir Annexe 2.

<sup>2</sup> Clément OURY, *Elections 2007 : Bilan de l'organisation du projet*, op. cit., p. 41. Reproduit avec l'aimable autorisation de Gildas Illien.

<sup>3</sup> Op. cit., p. 30.

## 3.3 Intérêt de la coopération

### Des bénéfices réciproques

Les bénéfices réciproques de la campagne d'archivage électoral pensée comme un « banc d'essai » par la BnF sont évidents et ils permettent d'envisager une coopération à long terme. D'une part, la BnF assure le soutien d'une solide infrastructure technique, en permanente évolution, que le partenariat avec la société américaine Internet Archive, au sein du Consortium IIPC, inscrit dans une dimension internationale et de recherche en matière d'archivage du web. La BnF possède aussi les espaces de stockage et la maîtrise de la haute technicité nécessaire à l'enregistrement et à la conservation de données numériques massives difficilement accessibles aux bibliothèques municipales en termes de coûts d'équipement et de moyens humains. D'autre part, comme le soulignait un conservateur de la BnF lors des Journées des Pôles associés de novembre 2007, les établissements membres du projet ont pu « sensibiliser [leurs] équipes [...] aux enjeux du numérique et de l'Internet en particulier. Cette constitution de fonds d'un nouveau genre s'inscrit aussi bien dans leur mission de dépôt légal que dans le cadre du recueil d'une documentation régionale et patrimoniale. Elle ouvre la voie à de nouvelles formes de valorisation des collections, complémentaires des documents conservés sur support. »<sup>1</sup>

### Valorisation de l'expérience au sein des BDLI

Plusieurs BDLI ont mis à profit l'archivage de la Web-campagne pour valoriser les ressources électroniques de l'internet en les insérant dans les collections traditionnelles. Nous avons déjà évoqué le cas de la bibliothèque de Lille qui a joint les liens des sites repérés à son catalogue et celui de Lyon où les sites électoraux des départements de Rhône-Alpes ont été intégrés aux dossiers du service « Points d'Actu ». Les lecteurs et le personnel sont par conséquent conduits à considérer ces liens comme un prolongement naturel des collections en direction de ce nouveau média de l'information qu'est l'internet. La bibliothèque acquiert par ce biais l'image d'un vecteur de l'actualité, à la fois expert en nouvelles technologies de l'information et médiateur entre le document numérique et ses lecteurs.

Si la loi limite pour l'instant la collecte du web et son archivage à la mission de dépôt légal de la BnF, cette expérimentation aura été l'occasion de faire appréhender aux bibliothécaires la complexité de l'objet numérique déposé sur l'internet. L'architecture d'un site web n'est pas simple : elle ne se donne pas d'emblée comme un objet cohérent. Elle se construit *a posteriori* à travers ses multiples liens. Peut-être ce projet aura-t-il eu aussi pour mérite de faire apercevoir l'articulation délicate entre ce qui ressortit aux missions traditionnelles du bibliothécaire, clairement identifiées (expertise du contenu, indexation, conservation) et ce qui renvoie à la spécificité du document numérique (notions de fréquence de captation, de profondeur, etc.). L'apport aura été à coup sûr méthodologique : un conservateur de BDLI<sup>2</sup> insiste sur la réelle « coopération scientifique » donnant lieu à une « politique documentaire qui se définissait à mesure ».

---

<sup>1</sup> Clément OURY in « Archiver les sites de l'Internet électoral de 2007 : une illustration à l'échelle nationale du dépôt légal de la Toile », 11<sup>e</sup> journées des Pôles associés, 22-23 novembre 2007, Strasbourg (document BnF non référencé).

<sup>2</sup> Bernard HUCHET (Bibliothèque municipale de Caen) dans l'enquête présentée à l'Annexe 1.

Il souligne également le « caractère empirique » de la recherche des sites, par-delà tout cadrage méthodologique et évoque le jeu complexe de « ricochets » et de « combines », la « patience » et « l'intuition » nécessaires à cet exercice. Un bénéfice *a minima* peut être retiré de l'expérience : elle aura contribué à représenter le site web comme un objet numérique apprivoisable par les bibliothécaires et apte à s'intégrer aux collections tant d'un point de vue documentaire que technique.

## **Perspectives**

Cette expérience de coopération peut s'avérer particulièrement profitable aux bibliothèques municipales soucieuses de prolonger leur fonds local par l'archivage du web. Dans l'enquête menée en préalable à cette étude<sup>1</sup> auprès des BDLI et des bibliothèques municipales, il apparaît que les principaux obstacles à ce projet sont de nature technique, financière, humaine (formation et mobilisation du personnel) et légale. Or, l'expérience menée en réseau semble apporter des réponses à ces interrogations. La BnF peut assurer le stockage et la conservation des documents numériques archivés en évitant, grâce au contrôle qualité, que des objets incomplets ou redondants encombrant des espaces coûteux.<sup>2</sup> Elle est en mesure d'adapter le dispositif à l'évolution des techniques en sollicitant au besoin des partenaires extérieurs. Le bilan souligne enfin qu'à l'issue d'une formation brève et à condition d'organiser un accompagnement des agents durant les premiers mois de repérage des sites, les compétences s'acquièrent d'autant plus vite que les critères auront été plus précisément établis, sans que cette fonction de veille et de préconisation mobilise un équivalent temps plein. Cette tâche, complémentaire d'activités bibliothéconomiques habituelles, peut requérir une journée par semaine par exemple, comme ce fut le cas pour certains bibliothécaires durant la campagne. Enfin, la restitution au public local peut s'opérer aisément, à condition que le décret d'application de la loi DADVSI étende les points de consultation, au moyen de la *Wayback Machine* d'Internet Archive par exemple, partenaire de la BnF et membre du Consortium IIPC.

## **4. ELABORATION D'UN SCENARIO POSSIBLE POUR UN RESEAU D'ARCHIVAGE DU WEB**

### **4.1 Une structure partenariale**

#### **Une collecte partagée**

Si l'on se fonde sur l'hypothèse selon laquelle un point de consultation des archives du web sera autorisé dans chaque région au sein des BDLI, il est possible d'envisager un scénario global d'archivage du web articulant les capacités techniques de la BnF et les compétences en matière de sélection et de préconisation des bibliothèques municipales. A la BnF reviendrait la centralisation de la gestion du réseau et les tâches largement automatisées, requérant une forte technicité, telles que le moissonnage, le traitement et

---

<sup>1</sup> Cf. Annexe 1.

<sup>2</sup> Une estimation croisant les coûts humains et techniques pour obtenir un coût « moyen » par unité technique de référence évalue à environ 88 euros l'acquisition le traitement et la conservation d'un Go. In Clément OURY, *Elections 2007 : Bilan de l'organisation du projet*, op. cit., p. 40.

l'indexation de la documentation numérique. Des bibliothèques en région émaneraient les propositions de capture saisies à travers un outil de sélection construit sur le modèle de celui qui fut mis en place lors de la campagne associée au web électoral, de manière à faciliter les recherches du robot moissonneur. Il resterait à la BnF à organiser les conditions d'un accès à distance dans chaque région, par exemple sous la forme d'une plateforme dédiée. Une réserve s'impose néanmoins dans le cas où des BDLI souhaiteraient une intégration plus forte afin de récupérer les données et de procéder de façon autonome à leur valorisation : quelle que soit la forme définitive que prendra le décret d'application, la loi ne permet pas une circulation des données, elle autorise seulement leur consultation.

## **La BDLI, échelon intermédiaire du réseau**

Dans la perspective d'une répartition des tâches entre la BnF et les bibliothèques en région, l'échelon intermédiaire et en quelque sorte « central », puisqu'il se situe à l'interface entre l'établissement national et les bibliothèques municipales, est celui de la BDLI. Ce positionnement dans le réseau lui confère un rôle à la fois centralisateur en matière de dépôt légal et fédérateur des demandes et des besoins qui émanent des bibliothèques partenaires. Bibliothèque de référence pour la documentation régionale, elle possède des fonds locaux et met en œuvre une politique d'acquisition qui en fait un interlocuteur constant et averti des établissements de la région, d'autant que « l'abandon progressif des anciennes provinces, la création des départements et des régions, celle des syndicats de communes et des districts [...] nuance considérablement la distinction fonds local/fonds régional. »<sup>1</sup> Yves Jocteur-Montrozier souligne à ce propos que le « fonds lyonnais » de la bibliothèque municipale de Lyon (BML) s'est mué en « fonds régional Rhône-Alpes ». De surcroît, ces établissements possèdent souvent une expérience en matière de référencement du web, voire de collaboration étroite avec la BnF dans l'archivage du web pour neuf d'entre eux, qui les placent naturellement dans la situation de référent régional. Ainsi, la BML a naguère constitué une base de 800 liens internet<sup>2</sup> dont nombre d'entre eux représentent un outil de travail appréciable en matière de documentation régionale.

A l'évidence, cette nouvelle fonction, qui rejoint les missions traditionnelles des BDLI, n'en nécessite pas moins des moyens supplémentaires, tant en formation du personnel qu'en moyens horaires pour assurer la coordination avec les principales bibliothèques de la région et avec la BnF tout en assumant une part de la sélection des sites. Dans la mesure où cette activité correspondrait à une extension de la mission de dépôt légal déjà déléguée en partie par la BnF, il semble possible de concevoir un financement sous la forme d'une subvention qui accompagne la charge nouvelle confiée aux BDLI. Il s'agirait en effet d'un effort d'archivage partagé dont les bénéfices reviendraient autant au fonds local des bibliothèques qu'à la BnF sous la forme d'une délimitation de corpus permettant de constituer des collections spécialisées au sein de l'archivage global de la Toile.

---

<sup>1</sup> Yves JOCTEUR-MONTROZIER, « Les fonds locaux » in *Le Patrimoine. Histoire, pratiques et perspectives*, dir. Jean-Paul ODDOS, Paris : Editions du Cercle de la Librairie, 1997, p. 356.

<sup>2</sup> Voir à ce sujet le travail d'Isabelle BONTEMPS, *Quelle politique documentaire pour l'acquisition de liens Internet en bibliothèque ?* Mémoire d'étude, Villeurbanne : ENSSIB, 1999.

## 4.2 Quels projets communs ?

### Exhaustivité ou sélectivité ?

Si ce schéma propose une répartition des tâches et indique les moyens d'une maîtrise des coûts, il ne résout pas l'alternative entre deux principes d'acquisition documentaire : la recherche de l'exhaustivité ou bien la sélectivité. D'emblée il nous faut apporter une nuance : si le volume des objets numériques sur le web et la volatilité des ressources électroniques rendent illusoire toute prétention à l'exhaustivité, du moins peut-on choisir de tendre vers une collecte aussi large que possible des ressources liées à la région sans critère discriminant autre que celui relatif au territoire. Par ailleurs, une bibliothèque municipale peut décider d'établir des critères de sélection rigoureusement définis, liés à la valeur du contenu documentaire des sites ou des bases de données, conformément à la politique d'acquisition de l'établissement : il s'agira alors d'un choix raisonné, parmi des ressources plus larges, fondé sur la recherche d'une cohérence avec la politique documentaire mise en œuvre pour les collections. Nous aurons l'occasion d'y revenir en observant le web forézien.<sup>1</sup>

La recherche de l'exhaustivité, pour légitime qu'elle soit dans les établissements ayant vocation à servir de pôle de référence régional (ainsi les BMVR), nous paraît peu adaptée aux établissements plus modestes, dont le rayon d'influence se limite par exemple au département. D'abord parce qu'elle nécessite une veille importante, consommatrice de temps, en ce qu'elle oblige à multiplier les tâtonnements, toute recherche s'avérant problématique dès que l'on sort des domaines clairement balisés par une identité locale forte par exemple. Cette activité est bien sûr possible, mais rapportée aux moyens dont disposent généralement ces établissements, et au regard de la valorisation assez limitée qui peut en être proposée en bibliothèque parmi les supports traditionnels, elle constitue une charge importante.

La structure collaborative coordonnée par la BnF au plan national et par les BDLI au niveau du maillage régional paraît plus adaptée à l'option de la sélectivité. Elle est le prolongement des campagnes de collecte du web électoral, mais elle n'implique pas pour autant une demande centralisée à la BnF ou à l'échelon régional. Des souhaits de collecte formulés par les bibliothèques municipales peuvent fort bien susciter un projet d'envergure beaucoup plus large ou à l'inverse trouver leur place dans un programme de dimension nationale. Ainsi, le principe de la collecte thématique semble offrir une assez grande souplesse, dès lors qu'une coïncidence peut être trouvée entre la volonté de donner une cohérence aux archives de l'internet amassées par la BnF dans le cadre du dépôt légal et la valorisation du web local en région. La capacité des bibliothèques à proposer des initiatives d'archivage est une condition importante de la réussite d'un tel projet collaboratif.

### Quelques pistes de projets

Signalons quelques exemples de projets fédérateurs. Une collecte ciblée autour d'une problématique sociétale telle que le développement durable permettrait un repérage du web militant, du web consacré par les entreprises à la promotion de cette dimension de leur activité, sans oublier les sites des administrations où l'action de l'Etat s'expose et

---

<sup>1</sup> Voir notre troisième partie.

incite à l'engagement. Un repérage des projets locaux devrait saisir le regard porté par les communautés sur l'aménagement du territoire : une forme d'expression nouvelle, moins tributaire du formalisme imposé par l'imprimé, est diffusée de manière spécifique par le web. A côté des collectes thématiques ponctuelles, un projet plus développé dans le temps pourrait faciliter le repérage des archives administratives en ligne dans le cadre du partenariat mis en place entre la BnF et les Archives nationales. Par le biais de conventions signées avec les administrations, les bibliothèques municipales pourraient recevoir le dépôt de ces sites et bases de données. Des projets ponctuels liés à des événements prévisibles (élections, manifestations cycliques de type local ou sportif) pourraient donner lieu à des collectes définies suivant un balisage territorial coordonné. La perspective d'offrir des corpus délimités aux chercheurs en sciences sociales s'avère une piste fructueuse : les communautés, les métiers, les sphères linguistiques constituent autant de domaines à même d'enrichir les collections locales et nationales consacrées aux territoires.

## **4.3 La place des bibliothèques municipales**

### **Le soutien des agences de coopération ?**

Dans le dispositif imaginé pour une collecte concertée, la BnF reste bien entendu l'établissement doté des moyens les plus importants dont il pourrait faire profiter les BDLI dans le cadre d'un élargissement de leur mission de dépôt légal au web. Aussi, la bibliothèque municipale semble-t-elle le maillon le plus vulnérable, dans la mesure où le temps consacré à la tâche ne fera probablement pas l'objet d'une rallonge budgétaire, le fonds local émergeant souvent pour une part très modeste au budget de l'établissement. On imagine mal comment la municipalité pourrait accorder un financement à un projet dont le retour en termes de ressources numériques et de valorisation des collections serait très faible, voire inexistant si l'on respecte une stricte application de la loi. Aussi peut-il être intéressant d'associer les Agences de coopération au réseau régional d'archivage de l'internet. Cette mission entre dans les attributions de ces organismes chargés de la promotion du livre et de la lecture à travers une programmation d'aides aux professionnels touchant à la valorisation culturelle, au conseil et au soutien économique notamment dans le cadre de projets liés aux nouvelles technologies. En Rhône-Alpes, nombre d'établissements ont bénéficié des aides de l'ARALD (Agence Rhône-Alpes pour le livre et la documentation) pour des chantiers tels que le microfilmage, la conservation partagée, ou bien encore la constitution de bases numériques par exemple. Par ailleurs, ce projet collaboratif s'inscrit dans l'objectif de mutualisation et de coordination à un niveau régional encouragé par les agences pour le livre. On peut concevoir leur rôle comme un appui aux bibliothèques municipales dans leur collaboration avec les BDLI : l'agence serait aussi un organe de discussion et d'échange facilitant la mise en place du réseau régional. Néanmoins elle ne se substitue pas à la fonction d'expertise de la BDLI qui demeure le garant scientifique du projet.

Cette intégration au projet serait d'autant plus naturelle que les « Pôles associés régionaux » de la BnF regroupent déjà au côté des agences de coopération les BDLI et les directions régionales de l'action culturelle (DRAC).

## Quel bénéfice pour les bibliothèques municipales ?

Grâce au soutien d'un dispositif cohérent mettant à leur disposition des instances de formation et de conseil ainsi que les prestations techniques de la BnF, les bibliothèques municipales pourraient envisager une collecte ciblée du web local, à la condition de disposer en retour des archives. Si un accès distant en BDLI est probable à moyen terme, l'extension à l'échelon de la bibliothèque municipale se heurte pour le moment aux dispositions juridiques. Seul l'établissement de conventions entre la bibliothèque et les producteurs de sites permettrait de recevoir un dépôt régulier de l'objet numérique, après chaque mise à jour éventuellement, ou bien autoriserait la capture dans des conditions conformes à la législation sur le droit d'auteur. Des initiatives ont déjà lieu en région : la Médiathèque de l'Agglomération troyenne (MAT), envisage la signature d'une convention avec l'Inspection Académique de l'Aube, afin de recevoir le fichier contenant le bulletin de cette administration. Les créateurs du web militant ou du web d'information à caractère régional ne devraient pas opposer de refus à cette tentative de valorisation des ressources locales qui est aussi la garantie d'un archivage de leur production. Les bibliothécaires peuvent aisément trouver des informations sur les noms de domaines existants par l'intermédiaire d'outils d'utilisation simple et rapide.<sup>1</sup> Cette solution *a minima*, en l'attente d'une ouverture du champ d'application de la loi, a le mérite de contribuer à créer une première collection limitée et représentative, issue pour partie des préconisations développées au sein d'un projet collaboratif, en mesure de valoriser à la bibliothèque le web local.

---

<sup>1</sup> Le moteur WHOIS, par exemple, délivre des informations sur la configuration du domaine, le serveur web et le propriétaire du site. Disponible à l'adresse <http://www.raynette.fr/services/whois/> [Consulté le 08/12/2008].

# Méthodologie pour une politique d'archivage

---

## 1. INTEGRER L'ARCHIVAGE DU WEB A LA POLITIQUE DOCUMENTAIRE DE L'ETABLISSEMENT

La constitution d'archives du web n'a de sens que par rapport à une collection existante, elle-même formée en partie par les hasards de l'histoire de la bibliothèque et développée grâce une politique d'acquisition cohérente. Or toute politique documentaire vise la satisfaction d'un public – usagers et non-fréquentants – dont la communauté qui finance et administre la bibliothèque est la représentation démocratiquement choisie. Peut-être n'est-il pas inutile de rappeler ce truisme, tant la réflexion sur la documentation régionale semble isoler ce champ au sein de la géographie documentaire de la bibliothèque pour en faire un objet autonome soumis à ses propres lois d'accroissement. En 2004, Henri-Jean Martin rappelait à quel point il serait hasardeux de séparer les fonds anciens des acquisitions les plus récentes par exemple, plaidant pour une « vision totale du livre » :

« [...] Il ne doit pas y avoir de différence conceptuelle entre les [fonds de prêt] et les [fonds de conservation]. Les frontières intellectuelles entre les fonds anciens, les collections des XIX<sup>ème</sup> et XX<sup>ème</sup> siècles et les acquis récents doivent donc être réduites au minimum : il n'y a qu'une culture. »<sup>1</sup>

Avant de se pencher sur les critères de sélection des ressources électroniques, il importe de rappeler les principales caractéristiques du fonds local dans lequel les archives du web sont appelées à s'intégrer. Poser la question en termes de spécificité du fonds, de rapport au patrimoine et de politique documentaire, c'est déjà appréhender les publics, usagers habituels ou projetés en fonction desquels s'opéreront les choix de contenus.

### 1.1 Le fonds local en bibliothèque municipale

« Élément fondateur », « colonne vertébrale pour les collections », c'est ainsi que Bertrand Calenge décrit la place éminente du fonds local, « unique fonds spécialisé dont la présence soit affirmée et commune pour toutes les bibliothèques municipales, et uniquement pour elles. »<sup>2</sup> En somme, le fonds local participe à l'identité de la bibliothèque municipale et son enrichissement est considéré comme l'une des missions fondamentales rappelée par la Charte des bibliothèques en ces termes : « Les bibliothèques municipales ou intercommunales doivent constituer et entretenir, en concertation avec les archives et les musées, un fonds d'intérêt local. »<sup>3</sup> Soulignons que la collecte du web local présente en effet un intérêt certain pour les archives – nous avons déjà évoqué le partenariat entre la BnF et les Archives de France au sujet du web administratif –, et à cet égard la réflexion concertée entre les Archives municipales de Lyon et le centre Rize, *Mémoires & Société* de Villeurbanne sur une mémoire numérique

---

<sup>1</sup> Henri-Jean MARTIN, « Une vision totale du livre », *BBF*, 2004, n° 5, p. 21.

<sup>2</sup> Bertrand CALENGE, *Les Politiques d'acquisition*, Paris : Editions du Cercle de la Librairie, 1997, p. 176.

<sup>3</sup> *Charte des bibliothèques*, adoptée par le Conseil supérieur des bibliothèques le 7 novembre 1991, Titre III, article 24.

préfigure peut-être ce que pourrait être un dispositif local de repérage et de conservation partagée. Le Rize est une structure originale qui réunit une médiathèque, les archives municipales et un dispositif muséal : il semble que la cohérence d'un « fonds local », en termes d'offre pour un lectorat large, appelle un regroupement de ce type, qu'il se présente sous une forme physique et novatrice comme à Villeurbanne, ou sous l'aspect plus virtuel d'un portail commun à différentes institutions. Une bibliothécaire d'Aix-en-Provence forme ce vœu en matière de référencement et d'accessibilité des archives du web local, en suggérant la création d'un portail régional.<sup>1</sup> Le souci de mettre à la disposition du public le plus large les ressources locales doit s'accompagner d'une recherche de lisibilité des collections et en faciliter l'accès puisque le support numérique élimine tout risque de dégradation des sources lié à la consultation. Aussi ne faut-il pas perdre de vue qu'à côté du fonds de prêt traditionnel, à visée encyclopédique, le fonds local constitue la raison d'être de la bibliothèque municipale dans lequel la « collectivité historique et vivante »<sup>2</sup> reconnaît son identité.

## Un domaine de référence

Le fonds local s'appuie le plus souvent sur un fonds antérieur, constitué de collections de périodiques, de monographies, voire de dépôts et de dons liés de manière plus ou moins étroite à la région : il suffit parfois qu'un auteur y soit né pour qu'il ait été décidé d'en collecter les œuvres. Une politique d'acquisition fondée sur trois critères de base – le sujet, l'auteur et l'éditeur/imprimeur<sup>3</sup> – permet son accroissement autour de tout ce qui touche à « la région, tous les documents qui en proviennent, les livres qui en parlent, les auteurs régionaux, les collectionneurs et leur bibliothèque. »<sup>4</sup> Bertrand Calenge distingue plus précisément quatre catégories :

- la documentation acquise pour le fonds en fonction des priorités définies,
- les documents dont l'intérêt local n'est que partiel (une simple copie pourra s'insérer dans un dossier),
- les documents collectés par furetage (tels que les tracts électoraux, les programmes culturels),
- et enfin ceux dont la conservation incombe à la bibliothèque par leur caractère local (enregistrement vidéo d'une manifestation locale récente,... ).<sup>5</sup>

A l'intérieur de ce cadre géographique, le bibliothécaire peut se sentir tenu à l'exhaustivité. Néanmoins, comme le souligne Yvette Weber, « si l'intérêt de collections régionales réside en grande partie dans l'exhaustivité, ce principe est, pour un établissement, une prétention irréaliste ».<sup>6</sup>

Dès lors se pose la question des choix à opérer dans la quantité foisonnante de productions qu'impose l'écrit et plus encore le gisement d'informations de tous ordres de l'internet. Si l'« encyclopédisme » est une dimension que l'on prête volontiers au fonds local, Yves Jocteur-Montrozier préfère quant à lui parler d'« éclectisme » ou de « variété » tout en notant la prédominance du découpage propre aux sciences humaines, « comme par le passé ».<sup>7</sup>

---

<sup>1</sup> Réponse au questionnaire présenté en Annexe 1.

<sup>2</sup> Bertrand CALENGE, *op. cit.*, p. 176.

<sup>3</sup> Marie-Agnès IBAR, *Constituer un fonds local : l'exemple de la Bibliothèque municipale de Versailles*, Mémoire de DCB, ENSSIB : Villeurbanne, 1999, p. 13.

<sup>4</sup> Guy PARGUEZ, cité par Marie-Agnès IBAR, *op. cit.*, p. 14.

<sup>5</sup> Bertrand CALENGE, *op. cit.*, p. 181.

<sup>6</sup> Yvette WEBER, « Les collections régionales à la Bibliothèque Municipale de Lyon », *BBF*, 1997, n° 2, p. 59.

<sup>7</sup> Yves JOCTEUR-MONTROZIER, « Les fonds locaux » in *Le Patrimoine. Histoire, pratiques et perspectives*, dir. Jean-Paul ODDOS, Paris : Editions du Cercle de la Librairie, 1997, p. 357.

Il nous semble que le principe de sélectivité applicable aux ressources issues du web ne nécessite pas de poser un cadre conceptuel nouveau : il est tout à la fois le prolongement d'une collection de référence pour la bibliothèque et pour une aire géographique, et son extension vers les nouvelles formes d'expression qui caractérisent ce nouveau média. Un exemple d'initiative locale suffit à montrer qu'il n'y a pas de rupture entre le fonds local et les possibilités d'accroissement offertes par le web. La Bibliothèque municipale de Brest, par l'intermédiaire de l'outil collaboratif WikiBrest, sollicite les contributions de Bretons invités à raconter leur propre histoire autour de sujets d'intérêt local. Leurs récits, qui s'inscrivent dans toute une partie du web que l'on nomme parfois le « web biographique », rejoignent de cette manière les ressources de la bibliothèque. Il va de soi que les forums locaux ou régionaux relevant d'initiatives individuelles à l'occasion de recherches mémorielles ou participant d'un désir de libre-expression sont des terrains naturels de collecte du web local. Ces documents pourraient figurer dans la quatrième catégorie définie par Bertrand Calenge. Par ailleurs, les programmes encapsulés dans les sites des institutions culturelles relèveraient de la troisième catégorie, le site lui-même pouvant prendre place dans la première, au titre de la division traditionnelle des « Arts » et de la « Culture ». Leur capture ne serait pas une simple duplication du support papier : outre la facilité de la collecte associée à un critère de fréquence, ces documents numériques intègrent souvent des vidéos présentant de courts extraits du spectacle ou des interviews du metteur-en-scène. Enfin, et pour ne se livrer qu'à un bref échantillonnage de la première catégorie, on ciblerait les sites et bases de données relatifs à la géographie locale, à l'histoire politique, industrielle ou sociale dont le contenu documentaire relève à l'évidence des « Sciences humaines ».

### **Patrimonialité du fonds local**

Outre le critère de contenu, déterminant pour la sélection des documents à intégrer aux collections, le bibliothécaire doit garder à l'esprit un caractère spécifique du fonds local : « les collections anciennes et régionales sont considérées d'office comme patrimoniales » rappelle Pierre Guinard.<sup>1</sup> Ainsi, tout document du fonds local est destiné par nature à devenir un document patrimonial, ce qui revient à dire qu'il est acquis avec un objectif de conservation à long terme qui le convertira au final en élément du fonds ancien. A ce stade de la réflexion, la problématique de la conservation des supports ne se pose pas directement pour la bibliothèque municipale, puisque la loi impose l'archivage à la BnF dans le cadre du dépôt légal. Néanmoins, les documents numériques acquis par convention avec les titulaires de la propriété intellectuelle pourront être conservés sur site : le support le moins onéreux et le plus facile d'accès reste actuellement le compact disc (CD). Mais deux difficultés sont loin d'être résolues : il reste difficile de garantir la pérennité du mode de consultation lié aux formats d'enregistrement et de prévoir la résistance du support à l'altération. On considère généralement que le CD gravé est sujet à une dégradation de son contenu à partir de cinq ans. La solution d'un archivage partagé dont les prestations techniques seraient assurées par la BnF constituerait la meilleure garantie d'une conservation pérenne en mesure de profiter des migrations de support inhérentes à l'évolution des techniques.

---

<sup>1</sup> Pierre GUINARD, « Pratiques patrimoniales de la bibliothèque municipale de Lyon », *BBF*, 1996, n°3, p. 37.

La valeur patrimoniale du fonds local détermine un autre aspect important de la collection d'objets numériques à constituer : elle doit s'offrir comme un corpus à l'intention des chercheurs. Aussi s'agit-il non seulement de considérer les besoins des chercheurs actuels mais aussi d'anticiper sur les champs d'investigation à venir. Tâche ardue et hasardeuse, qu'il est peut-être possible d'aborder avec sérénité si l'on considère, à l'instar de la BnF, que le rôle du bibliothécaire n'est pas seulement de prévoir la demande mais aussi, pour une part, de la susciter. Dans un champ aussi démesuré et aussi touffu que l'internet, en constante mutation, au travers duquel il reste difficile de délimiter des territoires, notion du reste étrangère à sa logique propre, notre bibliothécaire contemporain peut être saisi de vertige face au « maelström » que décrit plaisamment Bertrand Calenge dans son dernier ouvrage.<sup>1</sup> Aussi la constitution de corpus cohérents, conçus soit comme le prolongement des collections patrimoniales de la bibliothèque en suivant les axes de la politique documentaire, soit dans le cadre de dossiers documentaires élaborés pour le public le plus large et associant tous les supports, est une première piste. Dans une perspective plus large, la conception de corpus thématiques au sein du dépôt légal du web, à l'occasion de campagnes nationales, permettrait à la fois le repérage de ressources caractéristiques du web local et leur valorisation auprès d'un public de chercheurs. N'oublions pas que le web peut être considéré comme une ressource patrimoniale à un double titre : comme vecteur de contenus et comme objet d'étude à part entière dans la mesure où il est le lieu de naissance tout à la fois de formes de documentation à l'architecture innovante et de modes d'expression communautaires ou individuels inédits.

## Territorialité

Le logique du web pourrait contribuer par ailleurs à rendre désuète la notion de « territorialité » propre au fonds local. Un lieu commun contemporain figure par un amusant paradoxe – les « territoires du web » – la déterritorialisation à laquelle nous serions confrontés dès lors que nous entreprenons une navigation sur l'internet. Or, il semble que la délimitation géographique du fonds local ne soit pas la moindre difficulté à laquelle sont traditionnellement confrontés les bibliothécaires. A en croire André-Pierre Syren : « *Selon leur taille, les bibliothèques municipales définissent des surfaces plus variables : de la commune aux régions limitrophes en passant par l'arrondissement et le département. En fait, l'élément d'appréciation majeur réside dans le « limitrophe ». Les bassins de chacun des fonds épousent des formes étranges, rarement conformes aux découpages administratifs. Une cartographie détaillée permettrait assurément de reconnaître regroupements et chevauchements, les gros établissements constituant peu à peu des fiefs, les petits reconnaissant des suzerains. Le problème n'est d'ailleurs pas tant d'établir la légitimité de ces zones que d'en garantir la connaissance à autrui. D'illustrative, la cartographie deviendrait alors projective.* »<sup>2</sup>

En effet, au gré des découpages administratifs, l'aire couverte par le fonds « local » ou le fonds « régional » tendrait à se modifier, du moins sur ses frontières, la logique actuelle voulant que la bibliothèque de la préfecture s'attache au département, et la bibliothèque de la préfecture de région, souvent « à vocation régionale », à l'ensemble des départements de la province. Néanmoins, la présence de collections anciennes, à l'origine du fonds patrimonial et constituées suivant des découpages administratifs

---

<sup>1</sup> Voir la « Conclusion » in Bertrand CALENGE, *Bibliothèques et politiques documentaires à l'heure d'Internet*, Paris : Editions du Cercle de la Librairie, 2008, p. 257-260.

<sup>2</sup> André-Pierre SYREN, « Les fonds locaux de Champagnes-Ardennes », *Bulletin du CRL*, 1993.

désormais abandonnés, impose à la bibliothèque ses propres frontières. De surcroît, nombre de noms correspondant aux anciennes provinces ont presque disparu des cartes et des annuaires administratifs, mais ils n'en continuent pas moins à vivre une seconde existence à travers des noms d'associations, des titres de revues de sociétés savantes, voire des noms d'entreprises. A la dimension mémorielle d'un tel phénomène s'ajoute une valeur identitaire qui transcende les délimitations géographiques imposées par les réformes successives de l'Etat. Ainsi, le Forez, « ancienne province de France », selon l'encyclopédie, donne lieu à quelques 964 000 références sur l'un des moteurs de recherche les plus utilisés par les internautes.<sup>1</sup> Or son tracé ancien recouvre aujourd'hui trois parties de départements et chevauche deux régions. Aussi pouvons-nous souscrire à cet avis d'Yves Jocteur-Montrozier : « *Il est [...] évident que le territoire des fonds régionaux ne recouvre pas obligatoirement le découpage administratif mais que jouent également les zones économiques, culturelles, linguistiques, ainsi que, parfois, l'intensité des relations.* »<sup>2</sup> Et ce n'est pas le moindre paradoxe de l'internet, que de donner à voir des représentations du territoire qui superposent les délimitations actuelles et celles du passé, opérant à l'occasion une reterritorialisation des lieux dans leurs frontières anciennes. Les communautés qui s'exposent sur le web échappent souvent aux divisions administratives et composent une cartographie nouvelle en fonction de critères historiques, sociaux, économiques ou linguistiques. Aussi la notion de frontière s'avère-t-elle particulièrement poreuse dès lors qu'il s'agit de considérer des sources non institutionnelles dont l'extension se mesure davantage en termes d'influence que de périmètre géographique. Enfin, le web favorise l'inscription de l'action locale à l'intérieur de réseaux nationaux, voire internationaux. C'est pourquoi il peut être intéressant d'évaluer un site web « local » à l'aune d'autres régions, dans le cas d'une architecture d'envergure nationale (le web militant s'y prête particulièrement) : s'il représente pour le moment une « capsule vide », la veille consistera à déterminer le moment où il acquiert un contenu propre, révélateur d'initiatives locales, afin d'en décider la capture et l'archivage.

## 1.2 La politique documentaire à l'heure de l'internet

Le bibliothécaire soucieux de maintenir une politique d'acquisition cohérente dès lors qu'il cherche à intégrer les nouveaux « territoires » du web peut légitimement se demander s'il n'y a pas lieu de réévaluer la « politique documentaire » en fonction de son nouvel objet. C'est l'une des questions que se pose Bertrand Calenge dans son ouvrage paru récemment : *Bibliothèques et politiques documentaires à l'heure d'internet*.<sup>3</sup> Vers quelle évolution faut-il tendre de manière à prendre en compte l'«hétérogénéité mouvante des pratiques d'appropriation et d'échanges d'information, [la] multiplication des vecteurs d'information et de culture, [les] bouleversements technologiques et sociaux, [l'] émergence de multiples acteurs inédits, etc.»<sup>4</sup> ? La réponse qu'il propose ne semble pas imposer de bouleversements fondamentaux. Du côté du service au lecteur, il lui paraît préférable de substituer à la traditionnelle fonction de « médiation », depuis longtemps revendiquée, celle d'«accompagnement » car en la matière le bibliothécaire « n'est pas nécessairement plus savant »<sup>5</sup> que l'utilisateur.

---

<sup>1</sup> Sur *Google*, le 19/11/2008.

<sup>2</sup> Yves JOCTEUR-MONTROZIER, « Les fonds locaux » in *Le Patrimoine. Histoire, pratiques et perspectives*, dir. Jean-Paul ODDOS, Paris : Editions du Cercle de la Librairie, 1997, p. 356.

<sup>3</sup> Paris : Editions du Cercle de la Librairie, 2008.

<sup>4</sup> *Op. cit.*, p. 257.

<sup>5</sup> *Id.*, p. 360

Encore faut-il ajouter que pour ce qui touche à la consultation des archives du web cette fonction semble prépondérante dans la mesure où, pour l'instant, ces dernières se présentent fréquemment comme des documents « troués », lacunaires en plusieurs endroits. Par ailleurs, la « navigation » au sein d'archives ne s'effectue pas exactement comme une navigation sur le web, nombre de liens hypertextes n'étant plus valides. Du reste faut-il pouvoir choisir entre différentes saisies d'un même site : la prise en compte du critère de fréquence de la capture peut être un critère utile à considérer. A la fonction d'accompagnement, Bertrand Calenge ajoute trois dimensions essentielles pour déterminer la place de la politique documentaire dans le nouvel univers informationnel : il faut y intégrer la « dimension locale », la « mémoire vivante » et les « contenus de savoir ». Si l'on garde à l'esprit qu'il ne s'agit plus de la seule « mise en ordre raisonnée et prescriptive des collections »<sup>1</sup>, mais de l'invention de nouveaux usages permettant de donner forme et visibilité aux ressources de la bibliothèque grâce à des actions d'information, à des actions culturelles à destination du public le plus large, on mesure à quel point l'insertion d'archives du web – si l'on excepte les difficultés techniques de restitution, pour l'instant réelles – offre de réelles opportunités sans introduire de grande modification dans la politique documentaire conçue comme « l'art de gérer une bibliothèque au mieux de ses missions fondamentales. »<sup>2</sup>

Il est révélateur de constater que Gildas Illien inscrit lui aussi son approche dans la continuité des services traditionnels offerts par la bibliothèque. Dans sa contribution au Congrès de l'ADBU de septembre 2007, intitulée « Repenser la politique documentaire à l'échelle de la Toile. : l'expérience du dépôt légal d'Internet à la BnF »<sup>3</sup>, il adopte une perspective inverse de la tendance généralement constatée qui consiste à « déplacer la politique documentaire de la constitution de l'offre vers la valorisation des accès et services », le plus souvent hors de la bibliothèque, dans l'environnement internet de l'utilisateur. La délimitation de collections au sein des archives du web lui semble au contraire montrer « qu'il reste possible et utile d'intervenir en amont dans la constitution de l'offre ». Si la démarche d'acquisition doit tenir compte des contraintes liées à un nouveau support, « le pilotage des technologies de capture du Web offre un autre exemple où la responsabilité documentaire continue de s'exercer, mais autrement : à d'autres échelles, avec d'autres outils. » Aussi l'évolution du métier à envisager se pose-t-elle essentiellement en termes d'acquisition de compétences et d'organisation du travail en réseau, la spécialisation des tâches imposant un double impératif de partage des rôles et de formation continue. En somme, pour G. Illien, l'intégration des archives du web aux collections de la bibliothèque nécessite une adaptation qui prenne en compte différents facteurs dont certains ne s'appliquent pas seulement au dépôt légal du web mais aussi aux chantiers qui attendent les bibliothèques municipales soucieuses de capturer ces ressources :

- une connaissance satisfaisante de l'Internet, qui permette d'appréhender le document dans au moins trois dimensions : le contenu, le niveau logique et le niveau physique,
- une redéfinition des principes d'enrichissement des collections afin d'en délimiter les champs,
- l'invention de nouveaux modèles d'organisation et de coopération (au niveau régional et national),

---

<sup>1</sup> *Ibid.*, p. 260.

<sup>2</sup> *Ibid.*

<sup>3</sup> Disponible en format *pdf* à l'adresse suivante : [http://www.adbu.fr/IMG/pdf/Gildas\\_1.pdf](http://www.adbu.fr/IMG/pdf/Gildas_1.pdf) [Consulté le 19/11/2008]

– la restitution à l'utilisateur, sous un double aspect : quelles sont les contraintes liées à ces archives particulières, par nature incomplètes (« on archive un tissu de liens, pas des documents juxtaposés »), quels nouveaux champs d'analyse se découvrent ainsi au chercheur ?

Par-delà la nécessaire mutation des tâches liée à l'évolutivité des concepts nés de l'univers numérique, on mesure combien ce travail repose sur le savoir-faire bibliothéconomique acquis par les professionnels. La migration sur un nouveau support n'induit pas de modification fondamentale de l'approche documentaire. La décision d'archiver un blog, par exemple, devra répondre à des critères habituels touchant aux contenus, à la place de l'auteur ou bien de l'hébergeur dans la sphère publique, critères communs aux documents sur support physique. Tout au plus certaines données techniques à transmettre au robot moissonneur participeront-elles d'une analyse liée à la nature du site et à la démarche de constitution d'une collection : ainsi faudra-t-il déterminer, entre autres, la périodicité de la capture et le niveau de complétude souhaité. Néanmoins la complémentarité de la bibliothèque numérique avec les collections traditionnelles suffit à montrer qu'il n'y a pas substitution de supports, mais constitution d'un ensemble cohérent et diversifié. A ce titre, le Département des Arts et Spectacles de la BnF poursuit conjointement la collecte systématique des programmes des théâtres et la capture des sites de ces institutions culturelles, programmes et newsletters compris, car le constat s'impose que « le dépôt légal du Web est complémentaire du travail de collecte des éphémères mais ne s'y substitue pas. D'autant que les informations portées sur le programme et celles du site web ne sont pas systématiquement identiques.»<sup>1</sup>

Aussi le caractère particulier de la collection d'archives du web se situe-t-il peut-être davantage dans la nature des documents déposés au terme du processus que dans les opérations de sélection conformes aux procédures habituelles des bibliothèques. S'il est naturel de considérer ces archives comme un prolongement des collections de la bibliothèque, encore faut-il se demander quel en sera le public cible et ce qu'elles sont en mesure de lui apporter à côté des ressources physiques déjà bien connues.

### 1.3 Publics

Les publics visés par les archives du web comprennent d'une part les usagers habituels des fonds locaux, dans la mesure où cette documentation prend place au sein des collections déjà constituées, et de nouveaux publics que l'on peut supposer intéressés par les ressources électroniques en tant qu'objet spécifique.

Néanmoins la définition des publics fréquentant les fonds locaux et régionaux ne va pas de soi. Il faut tenir compte d'une évolution des usages analysée récemment par Laure Rioust : « le patient travail des érudits locaux a cédé en partie la place à une recherche loisir plus volatile et superficielle. »<sup>2</sup> Aussi, à la profondeur historique des collections attendue par l'utilisateur convient-il d'ajouter une ouverture plus large sur l'actualité et les centres d'intérêt de ce lectorat : « si beaucoup s'intéressent à leurs racines, leur manière d'accéder à l'histoire repose souvent sur un aspect ludique, émotif, ponctuel, et reste soumise aux modes. »<sup>3</sup> Aller à la rencontre des publics consiste dès lors à établir des

---

<sup>1</sup> Caroline RAYNAUD, BnF, Département des Arts du spectacle, courriel reçu le 07/11/2008.

<sup>2</sup> Laure RIOUST, *De l'érudit aux usagers. Publics des fonds locaux et régionaux en bibliothèque municipale : évolutions, pratiques et représentations*. Mémoire de DCB, ENSSIB : Villeurbanne, 2008, p. 85.

<sup>3</sup> *Idem*.

corrélations entre les sources documentaires, afin de parvenir à ce que des collections sur différents supports se valorisent mutuellement. Ainsi l'élaboration de dossiers documentaires à visée informative ou pédagogique, à la faveur d'événements marquants par exemple, sera en mesure de mettre en rapport les ressources traditionnelles et les archives du web. Leur présentation sur le site de l'établissement peut être l'occasion d'un signalement des nouvelles ressources et une invitation à leur consultation sur poste dédié à la bibliothèque.

Toutefois, il serait souhaitable que le bibliothécaire accompagne l'utilisateur dans sa consultation des documents numériques archivés tant leur caractère lacunaire peut être source de déception. En effet, le site web se présente comme une architecture déjà constituée, faite de liens hypertextes (dont certaines cibles n'auront pu être conservées) au sein de laquelle la recherche d'information peut s'avérer problématique. Les mots-descripteurs grâce auxquels un site a pu être indexé ne couvrent qu'imparfaitement le champ documentaire tant la recherche de l'information s'effectue naturellement sur l'internet par un jeu de « tâtonnements » et de « ricochets »<sup>1</sup>. De ce parcours interrompu dont l'archive n'offre que les instantanés de quelques étapes peut résulter un sentiment de perplexité. Gildas Illien rapporte cette expérience qui consista à fixer un thème de recherche parmi les archives du web électoral à deux juristes de formation comparable possédant une maîtrise équivalente de l'internet.<sup>2</sup> Il s'est agi en un temps limité de recueillir le plus grand nombre d'informations sur le sujet posé. Un expérimentateur s'avoua pleinement satisfait de la documentation mise à sa disposition tandis que l'autre exprima le jugement inverse. Aussi l'action de médiation s'impose-t-elle non seulement en ce qui concerne la présentation des archives de l'internet au public mais aussi en matière d'accompagnement dans la recherche de l'information.

Si la problématique de l'accès s'avère essentielle, la volonté de constituer une offre peut guider le sélectionneur en fonction des représentations que l'on est d'ores et déjà en mesure de se faire des publics intéressés par les archives du web. On peut légitimement s'attendre à rencontrer la demande des auteurs ou éditeurs ayant perdu trace de leur site internet ou de leur blog : un accès au dépôt légal du web pourra peut-être leur apporter satisfaction. Mais la documentation réunie touche au premier chef les chercheurs des sciences humaines, de l'information et de la communication pour qui l'internet est un objet d'étude au sein duquel il s'agit de délimiter des champs d'analyse. Une vision prospective peut s'appuyer sur les orientations actuelles de la recherche. Le travail de Dominique Cardon<sup>3</sup>, par exemple, chercheur associé à l'EHESS, est révélateur de nombre de problématiques tournées vers les « transformations contemporaines de l'espace public » induites en partie par les « infrastructures de médiation offertes par les nouvelles technologies ». Les « dynamiques d'individualisation et de participation à la production d'expression publique » sont observées par le sociologue sur de nombreux terrains parmi lesquels les outils d'expression individuelle sur le web que sont les blogs ou le wiki occupent une place privilégiée. Enfin, les « formes d'innovations par l'usage dans le développement des nouveaux médias de l'Internet » retiennent l'attention du chercheur.

---

<sup>1</sup> Nous empruntons ces termes à Bernard HUCHET, dans une réponse au questionnaire donné dans l'Annexe 1.

<sup>2</sup> Rapporté lors de l'entretien à la BnF du 28/10/2008.

<sup>3</sup> Voir la fiche de présentation de Dominique CARDON sur le site de l'EHESS : <http://cems.ehess.fr/document.php?id=155> [Consulté le 19/11/2008]

Dans le champ plus restreint du web local, on conçoit sans peine que des corpus puissent être délimités en privilégiant les centres d'intérêt de publics « spécialisés », qu'il s'agisse de la demande d'élèves ou d'étudiants souhaitant mesurer l'écho régional d'un événement particulier ou bien d'une collaboration avec le laboratoire d'une université proche. La capture ponctuelle et ciblée semble la procédure la plus adaptée à la constitution de ces corpus. Pour autant il ne faut pas négliger la mise en place d'une veille du web dit « sociétal », considéré comme un sismographe de l'opinion exprimée sur l'internet. Ainsi, l'attention aux publics potentiels nécessite une capacité d'anticipation et une certaine réactivité à l'événement afin de proposer une offre en mesure de valoriser les archives du web et de susciter l'intérêt du public, sans exclure la possibilité de partenariats dans le cadre de demandes spécifiques.

## 2. PARAMETRES A PRENDRE EN COMPTE

Si l'on excepte momentanément, dans le travail du sélectionneur, le critère de contenu qui sera déterminé en fonction de la politique documentaire de l'établissement et plus particulièrement de la charte des collections du fonds local, il est nécessaire de se pencher sur deux paramètres essentiels, demandés par le robot moissonneur, dont les implications ne sont pas seulement techniques. En effet, préconiser une *fréquence* (ou *périodicité*) de capture implique une observation attentive des sites : à cet égard une typologie (forcément incomplète) peut aider le sélectionneur. En outre, choisir le degré de complétude du site à collecter – c'est-à-dire déterminer la *profondeur* requise –, afin de ne cibler que les contenus pertinents, exige une connaissance minimale de l'architecture du web. Aussi peut-il être utile de s'interroger sur la notion d'« unité », qui est la base du référencement et de l'évaluation du volume d'une collection.

### 2.1 Quelle unité de collecte ?

Pour qui veut appréhender le web local, l'unité de référence qui semble s'imposer est le « site ». Néanmoins, le résultat d'une simple requête à travers un moteur de recherche laisse apparaître des unités éditoriales diverses : sites web, fichiers au format *pdf*, blogs, pages web contenant des fichiers de sons ou d'images encapsulés, etc. Cette simple observation donne la mesure de la révolution qui s'accomplit avec l'internet dans le domaine de la hiérarchisation des unités documentaires : si la logique de l'écrit, avec ses concepts de *page*, d'*auteur* et d'*éditeur* peut offrir des représentations qui permettent de se figurer certains objets numériques, elle n'est plus à même de rendre compte de l'architecture propre à l'internet. A la nature sérielle du support papier, caractérisée par la multiplication raisonnée d'unités de formats déterminés (monographies ou périodiques), s'est substituée sur le web une logique de l'emboîtement et de l'arborescence sans limite préétablie. Aussi la notion de *profondeur* correspond-elle précisément au chemin que le robot emprunte pour parcourir la Toile après y avoir accédé par l'intermédiaire d'une adresse. Cette adresse de départ étant qualifiée de « graine », on choisira le niveau de *granularité* souhaité dans la capture d'ensembles plus ou moins étendus, plus ou moins profonds.

Ainsi, pour l'échelle la plus large, il sera possible de s'intéresser à des *réseaux de sites amis*, renvoyant généralement à un contenu documentaire commun, et correspondant à la pratique exploratoire de l'internaute que ne limite pas l'unité du site. A un niveau de granularité inférieur, le *site* (et particulièrement les sites de grandes organisations) peut

lui-même héberger d'autres sites. La *page web* possède une identité propre qui se nourrit d'autres unités : *images*, *vidéos*, etc. Elle est un fichier qui contient d'autres fichiers encapsulés. Un répertoire d'images peut fournir de précieuses unités documentaires au sélectionneur dans la perspective d'un enrichissement du fonds iconographique local.

Le découpage technique du web peut d'ailleurs fournir à l'occasion des unités significatives, bien que son mode d'organisation ne se superpose pas à la hiérarchie documentaire que nous venons d'évoquer. Les domaines de haut niveau (*Top Level Domains*) sont répartis en deux catégories : les domaines *génériques* (en *.com*, *.org*, *.net...*) et les domaines *nationaux* (en *.fr*, par exemple, et en *.re* pour La Réunion). Or ces noms de domaines peuvent être achetés puis divisés par leurs propriétaires en de multiples noms d'*hôtes* eux-mêmes souvent subdivisés en *dossiers*. C'est notamment le cas des partis politiques qui acquièrent un nom de domaine dans lequel on trouvera autant de noms d'hôtes que de fédérations.

Retenons que la notion d'unité est délicate à définir dans le contexte de l'internet, et que la délimitation de l'extension à saisir devra s'appuyer dans chaque cas sur l'observation du contenu à archiver. En d'autres termes, la logique de la capture correspondra à une logique documentaire subordonnée à deux principes : l'intérêt du contenu pour les collections et son positionnement exact dans l'arborescence.

## 2.2 Le critère de profondeur

L'adresse indiquée au robot moissonneur fonctionne comme une porte d'entrée et ne constitue pas une cible de capture. C'est pourquoi la *profondeur* doit délimiter le nombre d'unités à collecter dans l'arborescence de l'objet numérique, en fonction des contenus documentaires visés. Aussi ce paramètre repose-t-il sur la compréhension d'une adresse URL (*Uniform Resource Locator*). Pour le robot, la page est l'unité la plus petite à collecter et le domaine représente la plus vaste. Une adresse sur l'internet répond à la structuration suivante en six éléments séparés par des signes de ponctuation :

`protocole_de_transfert://hôte.domaine.extension/chemin/page`

Différentes *profondeurs de capture* peuvent être envisagées :

- par **domaine** :

Le robot capture toutes les extensions du site (que l'on peut symboliser par une étoile : `http://*.domaine.fr/*`) : fichiers, serveurs hébergés,... Il est à noter que le nom de domaine correspond à la chaîne « **domaine.extension** ».

*exemples* :

- <http://www.forez-info.com/>
- <http://www.loireforez.fr/>

- par **hôte** (*host*) :

L'*hôte* est souvent un serveur, mais pas obligatoirement. Le robot capture tous les fichiers contenus dans un serveur hébergé sur un site à l'exclusion des autres. Le découpage en hôtes est utilisé par les plateformes d'hébergement de sites (*free.fr* par exemple) et de blogs (*blogspot.com* par exemple). Cette profondeur est donc adaptée à la captation de blogs. Des sites particulièrement volumineux ont aussi recours à ce découpage pour identifier plus facilement des unités de contenu : le site [www.saint-](http://www.saint-)

[etienne.fr](http://etienne.fr) correspond à la ville de Saint-Etienne ; <http://musees-scolaires.saint-etienne.fr> présente les musées de la ville dans la perspective d'animations pédagogiques.

*exemples :*

- <http://forezhistoire.free.fr/> (seules les pages du site « forezhistoire » sont captées, à l'exclusion des autres pages de « free.fr »)
- <http://othl.over-blog.com/> (sur « over-blog.com », seul le blog correspondant à « othl » – Office de tourisme du Haut Livradois – est capturé)

- par **chemin** (*path*) :

On appelle *chemin* la succession des répertoires en partant d'une racine jusqu'à un ou plusieurs fichiers cibles à l'exclusion des autres. Cette profondeur correspond à un niveau de l'arborescence du site dans lequel est rangé un ensemble cohérent de documents. Certaines plateformes d'hébergement utilisent les chemins pour les sites qu'elles hébergent (*orange.fr*, *chez.com*, *lycos.fr*)

*exemple :*

- <http://www.over-blog.com/recherche/blogs/forez>  
(Cette profondeur permet de capturer sur « over-blog.com » les blogs ayant un lien avec le Forez)

- par **page** et **contexte** :

Le contexte n'est pas un élément de la structure d'une adresse internet, mais c'est une indication qui impose au robot de collecter avec la page (y compris ses animations, ses images, ses vidéos), toutes les pages qui se trouvent à un clic au-delà de la page et qui font partie du même domaine. Cette profondeur est particulièrement utile pour la presse en ligne par exemple.

*exemple :*

- <http://forezhistoire.free.fr/vie-religieuse.html> (la page de présentation de la rubrique « vie religieuse » du site « forezhistoire » est capturée ainsi que toutes les pages de la rubrique à un clic de distance, ce qui permet de disposer de la rubrique au complet)

- par **page** :

Une seule page est capturée : cette profondeur est adaptée aux captations fréquentes en raison de la légèreté de la manipulation

*exemple :*

- <http://forezhistoire.free.fr/sainte-claire.html> (seule la page consacrée à « sainte-claire » est capturée au sein de la rubrique « vie religieuse » du site « forezhistoire »)

Le critère de profondeur sera déterminé en fonction de l'extension du champ documentaire retenu pour la collecte sur l'arborescence. Il va de soi que le site de la municipalité de Saint-Etienne, par exemple, mérite une capture globale dans la mesure où il héberge de nombreux services : ceux de la municipalité, les archives de la ville, le site de la médiathèque, etc. Une observation attentive de ces unités plus ou moins volumineuses et englobantes doit permettre à la fois d'éviter les « doublons » dans les captures et de pallier les lacunes de la collecte. Pour reprendre l'exemple précédent, si les musées de la ville sont valorisés et amplement présentés sur le site [www.saint-](http://www.saint-)

[etienne.fr](http://www.mam-st-etienne.fr), les musées dont l'aire culturelle dépasse la région possèdent souvent leur propre site. Ainsi le Musée d'Art moderne de Saint-Etienne à l'adresse <http://www.mam-st-etienne.fr>.

## Le web invisible

Il est un obstacle bien connu à la recherche du robot en profondeur, c'est le web dit « invisible » parce que son indexation n'est pas reconnue par les moteurs de recherche traditionnels. Son volume est considéré comme bien supérieur à celui du web « visible ». Des sources documentaires très diverses sont concernées, qu'il s'agisse d'informations à renouvellement très fréquent (les horaires de chemin de fer, par exemple), de l'information spécialisée (la « littérature grise » notamment), ou bien encore de bases de données accessibles sur inscription et à l'aide d'un mot de passe. La seule attitude pragmatique que l'on puisse adopter consiste à évaluer l'intérêt potentiel de ces objets numériques et à solliciter du propriétaire du site l'autorisation d'y accéder ou bien à sonder ce champ à l'aide de différents instruments désormais disponibles si l'on a lieu de croire que certains domaines de la recherche possèdent une extension dans le web invisible.<sup>1</sup>

### 2.3 Le critère de fréquence

La périodicité de la capture est elle aussi subordonnée à la spécificité du document à archiver. Il est essentiel de déterminer la fréquence du renouvellement de l'information. Encore faut-il estimer la « surface » de ce renouvellement. Dans la perspective d'une consultation des fonds issus de l'internet, il est souhaitable d'éviter une trop large redondance du contenu dans les versions d'une même unité capturées successivement. Les sites ne sont pas comparables à des périodiques et leur mise à jour ne correspond le plus souvent qu'à des modifications ou apports partiels. Par ailleurs, nombre d'entre eux intègrent une fonction d'auto-archivage (c'est le cas des blogs notamment). Aussi peut-on proposer, à titre indicatif, quelques fréquences de capture en fonction d'une typologie sommaire des sites, l'analyse la plus adaptée relevant du cas par cas.

<i>Types de sites</i>	<i>Fréquences préconisées</i>	<i>Exemples</i>
• Sites avec renouvellement rapide sans auto-archivage ( <i>sites de presse, d'information générale,...</i> )	1 capture par semaine	<a href="http://www.forez-info.com">www.forez-info.com</a> site très bien documenté accordant une large part à l'actualité

<sup>1</sup> Cette page de l'université de Nantes, par exemple, répertorie plusieurs outils et métamoteurs utiles pour effectuer une recherche dans le web invisible : <http://www.sciences.univ-nantes.fr/info/perso/permanents/desmontils/InvisibleWeb.html> [Consulté le 08/12/2008]

<ul style="list-style-type: none"> <li>• Sites avec renouvellement partiel et périodique (<i>sites des institutions culturelles, administratives, etc.</i>)</li> <li>• Sites avec renouvellement fréquent et auto-archivage (<i>blogs,...</i>) ; web collaboratif</li> </ul>	1 fois par mois	<a href="http://www.comedie-de-saint-etienne.fr">http://www.comedie-de-saint-etienne.fr</a> site de la Comédie de Saint-Etienne <a href="http://portail.univ-st-etienne.fr">http://portail.univ-st-etienne.fr</a> portail de l'Université Jean-Monnet de Saint-Etienne <a href="http://saintchamond.canalblog.com">http://saintchamond.canalblog.com</a> blog consacré à l'histoire et à l'actualité de Saint-Chamond (base photographique importante)
<ul style="list-style-type: none"> <li>• Sites à caractère érudit, stables (information de nature historique renouvelée par des ajouts occasionnels, par exemple), et possédant un auto-archivage</li> </ul>	tous les 2 mois	<a href="http://www.ladiana.com">http://www.ladiana.com</a> site de la Société historique et archéologique du Forez <a href="http://www.rene-merle.com">http://www.rene-merle.com</a> site d'un érudit dont les pages consacrées à la recherche linguistique (occitan et francoprovençal) présentent un grand intérêt <a href="http://www.arpitania.eu/">http://www.arpitania.eu/</a> site sur l'arpitan et partiellement en arpitan, stable et à enrichissement lent

<ul style="list-style-type: none"> <li>• Sites occasionnels (liés à un événement et dont la durée de vie est limitée)</li> </ul>	au moins 1 capture  à définir au cas par cas ( <i>avant et après une date,...</i> )	- le web électoral - les blogs destinés à accueillir une participation citoyenne à l'occasion d'un grand projet : <a href="http://blog.parc-livradois-forez.org/index.php?Developpement-durable">http://blog.parc-livradois-forez.org/index.php?Developpement-durable</a> (« atelier citoyen » autour d'un nouveau projet pour le parc Livradois-Forez). Dans le cas d'un projet à long terme comme celui-ci, une <i>périodicité plus rapprochée</i> est souhaitable : pour un blog avec auto-archivage, 1 à 2 fois par mois.
--	--	---

Par la notion de *stabilité*, nous entendons la permanence des sites sur le web, liée soit à leur caractère institutionnel (web administratif, mais aussi de sociétés savantes anciennes et reconnues), soit à la persévérance de leur auteur, qu'il s'agisse d'un groupe de passionnés ou de militants, ou bien encore d'un individu isolé attaché à la valeur de l'information qu'il rend publique. Certains sites d'érudits ont une durée de vie plus longue que des sites d'établissements scolaires ! Les publications périodiques attachées à certains sites (qui se substituent parfois à la version papier) peuvent faire l'objet d'une collecte distincte ou bien déterminer la fréquence de capture du site.

Pour les sites dont le critère de périodicité est difficile à évaluer, parce qu'ils ne semblent pas se rattacher aux catégories générales pour lesquelles la procédure ne pose guère question, il est bien sûr utile de rechercher la date d'actualisation, mais il est aussi très suggestif de comparer plusieurs versions archivées en utilisant la *Wayback Machine* de Internet Archive.<sup>1</sup> De cette manière, on pourra vérifier la longévité du site sur l'internet et estimer la périodicité et l'ampleur de l'enrichissement effectué par les auteurs.

### 3. ELABORATION D'UNE GRILLE DE SELECTION

Les critères qui précèdent, requis par le robot moissonneur, mais dont les différentes options relèvent d'un choix de nature documentaire, méritent d'être intégrés à une grille qui rassemble et formalise les critères de sélection d'un site. De nombreuses grilles d'évaluation des sites ont été proposées et il a même été recommandé de distinguer, pour la création d'une base de liens en bibliothèque, une première grille d'observation qui permette d'éliminer les sites non indiqués (à cause du format, de l'accès, de la langue, ou de la source) avant l'application d'une grille de sélection plus rigoureuse.<sup>2</sup> Notre conviction est que le suivi le plus efficace d'un corpus raisonné de sites web doit s'articuler autour d'un nombre de critères restreint. Le choix de s'intéresser au web régional constitue une première délimitation du champ documentaire. Par ailleurs, des projets plus ciblés, limités dans le temps ou nécessitant un suivi particulier (pour des problématiques régionales spécifiques par exemple) peuvent motiver des choix qu'une politique documentaire globale ne justifie pas. Dans la mesure où il est difficile de séparer critères techniques et choix documentaires, nous proposons d'intégrer ces données en un seul tableau. Trois champs peuvent rendre compte des informations nécessaires à la collecte et au suivi des sites sélectionnés.

#### 3.1 Les champs

##### Identification et validation documentaire

Le premier champ, consacré à l'«*identification du document numérique*» et à la «*validation documentaire*», s'attachera à caractériser le document, en essayant notamment de délimiter sa surface dans l'arborescence, en rapport avec le motif de la sélection qui fera l'objet d'une explicitation succincte.

Il conviendrait tout d'abord d'attribuer un nom au site, différent de l'adresse URL parfois peu explicite, de manière à le rendre lisible et repérable dans un système de référencement. A la date de consultation, il est souhaitable d'ajouter l'adresse URL correspondant à la plus large extension signifiante ; la profondeur de saisie retenue pour la capture sera précisée plus loin. Le repérage de la source ainsi qu'une brève évaluation de sa fiabilité constituent une première évaluation documentaire. Ainsi, un site administratif ou institutionnel possédera par nature un haut degré de fiabilité de l'information, mais un site personnel nécessitera une recherche plus approfondie sur son auteur, ou, à défaut, une brève évaluation de la qualité du contenu. Dans le cas de blogs

---

<sup>1</sup> Accessible à l'adresse <http://www.archive.org/web/web.php> [Consulté le 30/11/2008]

<sup>2</sup> Isabelle BONTEMPS, *Quelle politique documentaire pour l'acquisition de liens Internet en bibliothèque ?*, Mémoire de DCB, Enssib : Villeurbanne, 1999, p. 31 sq.

individuels ou de sites pour lesquels la qualité « scientifique » du contenu ne constitue pas le motif de la sélection, ce critère perd sa pertinence. Il serait utile de mentionner un contact, lorsqu'il existe, de manière à permettre l'élaboration d'une convention d'archivage avec une bibliothèque, par exemple, ou bien pour accéder au web invisible lorsqu'il représente une part intéressante et non confidentielle du document archivé.

Une évaluation sommaire du critère de validation documentaire constituerait une seconde partie de ce champ. Deux options s'offrent au bibliothécaire, correspondant à différents types de collectes. L'enrichissement du fonds local, d'une part, doit être motivé par la mention d'une (ou plusieurs) catégorie(s) préexistante(s) dans les collections (« Linguistique » dans l'exemple de notre grille) au besoin précisée (« Etude sociolinguistique et historique du francoprovençal » dans le même exemple), dans laquelle le document a vocation à s'insérer. Le choix d'une collecte ciblée, d'autre part, présente au moins deux possibilités : une collecte ciblée à long terme, de manière à repérer les sites caractéristiques d'un champ régional spécifique, ou bien une collecte ponctuelle destinée à capturer un ensemble documentaire représentatif d'une problématique : les sites relatifs à une manifestation prévisible ou consécutifs à un événement possédant un retentissement local, par exemple. La « situation du contenu dans l'arborescence » suggère une délimitation de la surface documentaire à archiver : son corrolaire est bien sûr le critère de profondeur destiné à la capture et précisé dans le « masque de saisie ». Bien que l'architecture d'un site soit sujette à modification, cette restriction du champ de capture est souhaitable dans la mesure où les coûts liés au stockage d'un objet en croissance constante imposent le discernement. Dans le cadre de la constitution d'une collection limitée, produit d'une sélection, une expertise documentaire « en amont » est nécessaire.<sup>1</sup>

Enfin, une rapide estimation du nombre et de la pertinence des liens proposés par le site est une indication appréciable dans le cas d'une extension de la recherche à d'autres ressources numériques se rattachant au même champ documentaire, qu'elle émane du bibliothécaire ou du lecteur, l'utilisateur se limitant rarement à la consultation d'un seul site. Une mention des fichiers particuliers encapsulés (photos, vidéos, bases de données,...) fournira une information tant documentaire que technique utile (certains sites d'« histoire locale » présentent des bases photographiques étendues qui en font toute la richesse).

## **Masque de saisie**

Le deuxième champ de notre grille réunit les critères utiles au robot moissonneur et induits par l'évaluation documentaire qui précède. C'est un « *masque de saisie* » qui précise l'adresse URL retenue, la fréquence et la profondeur de la capture. Dans le cas d'une prestation assurée par la BnF, le risque de doublonnage est désormais anticipé par des procédures automatisées qui se fondent sur la syntaxe de l'adresse. Nous proposons de rajouter une évaluation de la « stabilité » du site (zone grisée car ne désignant pas une consigne pour le robot) destinée au sélectionneur et au suivi du site : seront considérés comme « à surveiller » les sites dont la pérennité est difficile à prévoir. Le critère « occasionnel » caractérise les sites à caractères événementiel, dont la durée de vie est prévisible, notamment ceux qui font l'objet d'une collecte ciblée et ponctuelle.

---

<sup>1</sup> Il en va tout autrement de la logique du dépôt légal de l'Internet. Gildas ILLIEN estime que « les problématiques du volume et de l'éphémère convergent ainsi vers le déplacement des traitements scientifiques et humains, les plus longs et les plus chers, de l'amont vers l'aval de la gestion de la collection, afin de limiter les coûts ainsi que le risque d'une disparition pure et simple des contenus » in « Le dépôt légal de l'internet en pratique : les moissonneurs du web », *BBF*, 2008, n°6, p. 22.

## Indications de suivi

Le dernier champ, consacré au « *suivi* », est destiné à enregistrer la veille du site, en particulier si sa stabilité était difficile à prévoir. Il peut aussi mentionner les changements d'orientation dans la politique d'archivage du web. Il va de soi qu'il présente une moindre importance que les champs précédents. Il doit concerner un nombre limité de sites mais peut s'avérer fort utile dans le cadre de la transmission du travail du sélectionneur afin d'en assurer la continuité. Les formats susceptibles de poser des problèmes au robot peuvent être signalés.

Afin d'alléger le traitement manuel dans les phases de sélection et de saisie des données techniques, les rubriques dont le renseignement est impératif ont été signalées à l'aide d'un astérisque. Il va de soi que l'analyse de nombreux sites, dont la source est clairement identifiable, ne requiert qu'un repérage rapide, centré sur les informations à transmettre au robot. En définitive, cette grille se veut un instrument pratique de repérage et de justification dans le cadre d'un archivage limité, répondant à des choix documentaires clairement définis, et soucieux du temps humain et des coûts matériels investis dans cette tâche. Elle constitue néanmoins un document de référence pour la tutelle, complémentaire du projet d'archivage élaboré par la bibliothèque. Elle est aussi le témoin d'une démarche analytique dont il importe de garder la trace dans le cadre d'un partenariat avec d'autres établissements.

## 3.2 Proposition de grille de sélection

### I. Identification du document numérique et validation documentaire

<b>Nom attribué au site*</b>		<b>Date de consultation*</b>	
Site personnel de René Merle		30/11/2008	
<b>Adresse du site*</b>			
<a href="http://www.rene-merle.com">http://www.rene-merle.com</a>			
<b>Source*</b>			
René Merle, agrégé et docteur en histoire		<b>Fiabilité</b>	
<b>Contact possible</b>		<input checked="" type="checkbox"/> haute	
Courriel : contact@rene-merle.com		<input type="checkbox"/> moyenne	
		<input type="checkbox"/> faible	
<b>Contenu en rapport avec...*</b>			
 <b>le fonds local</b>		 <b>une collecte ciblée</b>	
<ul style="list-style-type: none"> <li>• LINGUISTIQUE</li> <li>Etude sociolinguistique et historique du francoprovençal</li> </ul>		<input type="checkbox"/> champ régional spécifique : _____	
		<input type="checkbox"/> projet spécifique / collecte ponctuelle _____	
		<input type="checkbox"/> autre : _____	
<b>Qualité du contenu documentaire*</b>			
<input checked="" type="checkbox"/> spécialisé <input type="checkbox"/> courant <input type="checkbox"/> vulgarisation			
<b>Situation du contenu dans l'arborescence</b>			
Rubrique « L'écriture en francoprovençal : Forez et Lyonnais, de la Révolution à la Monarchie de Juillet ». Page avec liens vers des articles connexes.			
<b>Liens</b>			
Nombreux liens vers des articles connexes du même auteur.			
<b>Fichiers particuliers encapsulés</b>			

### II. Masque de saisie pour le robot moissonneur

<b>Stabilité du site</b>	<b>Profondeur*</b>	<b>Fréquence*</b>
<input checked="" type="checkbox"/> forte <input type="checkbox"/> difficile à prévoir <input type="checkbox"/> site occasionnel	<input type="checkbox"/> domaine <input type="checkbox"/> hôte <input checked="" type="checkbox"/> page et contexte <input type="checkbox"/> page <input type="checkbox"/> chemin	<input type="checkbox"/> 1 fois par semaine <input type="checkbox"/> 1 fois par mois <input checked="" type="checkbox"/> tous les 2 mois <input type="checkbox"/> capture unique
<b>Adresse URL à saisir*</b>		<b>Date limite (collecte ponctuelle)</b>
<a href="http://www.rene-merle.com/article.php3?id_article=23">http://www.rene-merle.com/article.php3?id_article=23</a>		__ / __ / 20__
<b>Formats particuliers</b>		

### III. Suivi

<b>Nouvelles consultations</b>	<b>Observations / Décisions</b>



# Archiver le web forézien

---

## 1. EN AMONT, DEUX ORIENTATIONS POSSIBLES

Les considérations qui précèdent sur l'élaboration d'une méthodologie préalable à l'archivage du web régional ne sauraient tenir lieu de réflexion sur l'objet même de la collecte, objet au demeurant mouvant et aux contours variables selon l'orientation choisie par le sélectionneur. Dans le cas d'un projet tendant à l'exhaustivité, la représentation que l'on peut s'en faire est relativement simple : tous les sites produits « dans » ou « sur » une aire géographique. Néanmoins, on conçoit que cette perspective nécessite la mise en place d'une veille approfondie et régulière tant il est de la nature même du web d'être en constante évolution. Le recours à une procédure automatisée – à l'aide d'un robot *crawler*, par exemple – ne se substituera jamais de manière satisfaisante à une analyse documentaire, seule en mesure d'évaluer la pertinence des captures et la dissémination des ressources régionales sur le web. On comprendra aisément que des établissements plus modestes, dont la vocation régionale est moins marquée qu'une BMVR par exemple, fassent le choix d'une sélection « raisonnée », plus limitée en volume et volontiers centrée sur des sites institutionnels stables et sur des objets présentant une certaine qualité documentaire, comparables aux imprimés des collections et d'ailleurs pensés sur le mode de la transposition des supports de l'écrit vers le numérique.

Il nous semble cependant que ces deux modèles théoriques méritent d'être reconsidérés, à l'épreuve des faits. Dès lors qu'elle échappe à la démarche de dépôt légal, la « tentation de l'exhaustivité », appliquée au contexte de la bibliothèque municipale, suscite plusieurs interrogations. Par ailleurs, l'option de la « collection raisonnée » souffre précisément de l'image qui lui préexiste, celle des collections d'imprimés, et de ce fait risque fort de négliger la spécificité du document issu de l'internet. Autant dire qu'elle peut s'en trouver appauvrie, au demeurant moins dans l'esprit des sélectionneurs actuels que pour les usagers futurs, familiers de l'univers numérique dès l'âge des premiers apprentissages.

La commande de préconisations pour l'archivage du web forézien, à destination de la médiathèque de Saint-Etienne, nous fournit l'opportunité d'une réflexion sur le champ à délimiter dans l'internet, au sujet d'une région dont l'ancien nom déborde les frontières administratives actuelles, à l'image du web qui se joue des découpages historiques et géographiques.

Notre souci sera de proposer une réponse cohérente au scepticisme que nous avons parfois lu en dépouillant l'enquête préalable à cette recherche et retrouvé lors de nos entretiens, et que l'on peut résumer en une question simple et légitime : « à quoi bon fournir l'effort d'un long travail de sélection des sites à l'heure où la BnF réalise des captures larges de tout le domaine français ? »

## 1.1 La tentation de l'exhaustivité

### Les collections

Parmi les réponses à notre enquête<sup>1</sup> sur l'intérêt d'archiver le web régional, une seule préconise, à terme, la collecte quasi exhaustive des sites produits dans la région ou portant sur elle. Elle émane du responsable de la documentation normande et du patrimoine à la bibliothèque de Caen, Bernard Huchet, et situe la réflexion dans la perspective de l'ouverture d'une BMVR. Dans la mesure où les « circuits de l'information régionale » ont de plus en plus recours à la forme numérique, pour des raisons de coût et de diffusion, l'archivage du web local apparaît comme une nécessité afin d'assurer la « continuité d'une politique documentaire exhaustive » en matière de documentation régionale. Il s'agit dans un premier temps de prendre en compte la migration des supports afin de sauvegarder les informations diffusées sur les sites d'institutions, de sociétés savantes, de syndicats, de structures associatives dont les responsables n'ont pas toujours prévu l'archivage. Ainsi, le Jardin des Plantes de Caen publie désormais son *Index seminum* sous forme exclusivement électronique avec un renouvellement annuel. La bibliothèque serait donc en mesure de garantir la consultation rétrospective de « publications » de l'internet, par nature éphémères et soumises aux aléas des structures éditoriales de ce média : « les archives d'un blog ne sont disponibles que pour autant que le blog est disponible chez son hébergeur. » Par ailleurs, l'objectif reste d'anticiper une évolution qui encourage la publication sur l'internet : nombre d'auteurs, individuels ou représentant une collectivité, y gagnent une visibilité pour un faible coût. Dans le département de la Loire, 85 des 327 communes se sont dotées d'un site, sans que cette démarche soit en rapport avec la concentration de la population ; l'internet apparaît comme le moyen de valoriser les communautés qui ne possédaient pas de relais imprimé de l'information aussi bien que l'expression individuelle et la vie associative. C'est donc une nouvelle cartographie régionale qui se découvre sur le web, et que la bibliothèque de référence se doit de prendre en compte, dans un souci d'enrichissement et d'actualisation du fonds local. Ce constat amène deux conséquences : à terme, le projet doit tendre vers l'exhaustivité des sites produits en région et à propos du territoire, sans préoccupation de sélection documentaire, dans l'esprit d'une collection de dépôt légal numérique régional ; les bibliothèques titulaires du dépôt légal imprimeur apparaissent comme les mieux à même d'assurer cette mission en raison de « leur connaissance approfondie des circuits de l'édition régionale dont nombre d'acteurs sont également ceux du développement de l'internet ».

Cette manière d'envisager l'archivage du web local requiert un travail en réseau auquel répond particulièrement bien l'architecture de partenariat entre les bibliothèques et la BnF que nous envisagions dans notre première partie. Les contraintes de la loi DADVSI, les volumes de stockage, la nécessité d'un traitement largement automatisé des tâches en aval du processus de sélection et un coût difficile à assumer pour une collectivité locale plaident pour une collaboration étroite avec la bibliothèque nationale qui trouverait dans cet échange l'opportunité de délimiter des fonds régionaux cohérents au sein des archives du web, fonds aisément valorisables en région, voire à un niveau national. En outre, les préconisations issues des bibliothèques permettraient d'affiner la capture de certains sites en proposant une fréquence adaptée à leur objet, tenant compte du

---

<sup>1</sup> Cf. Annexe 1. Les citations de ce paragraphe sont de Bernard Huchet.

renouvellement de l'information. Ce projet pourrait mettre à profit la structure élaborée lors des captures du web électoral et s'y insérer moyennant quelques ajustements.

## Les territoires du web

Il nous semble par ailleurs que ce principe de capture offre un champ d'étude conforme à la nature du web. Les unités sélectionnées n'y apparaissent plus isolément, mais dans un réseau de liens dont on peut penser qu'il sera relativement complet pour les sites à problématique régionale. Il réserve surtout la possibilité d'étudier les relations entre un territoire à une époque donnée et la « couverture numérique », en termes de réseau informationnel et de contenu documentaire, qui s'y développe. De cette manière sont données à lire des dynamiques territoriales qui constituent un élément d'enquête fécond pour qui veut apprécier l'évolution d'un territoire.

Une étude menée en Picardie par le RENUPI<sup>1</sup> (Repères numériques de Picardie – Observatoire des technologies de l'information sur la région Picardie) en 2005 s'est fixée pour tâche de connaître la présence, la taille et l'organisation des acteurs présents sur le web. La récolte des données s'est effectuée à l'aide d'un robot *crawler* chargé de repérer à partir de quelques adresses les sites (nœuds) qu'il trouve en visitant chaque page, les liens (arcs) entre ces nœuds et le contenu sémantique des pages. 20 000 sites ont ainsi pu être répertoriés à partir d'une vingtaine d'adresses et ils ont fait l'objet d'un traitement manuel qui les a classifiés en fonction de trois critères : la localité, le type d'acteur (associations, entreprises, collectivités,...) et le thème (« tourisme et sport » par exemple). Les résultats obtenus ont fourni différentes indications : la taille permet d'évaluer la présence des acteurs, l'analyse des liens traduit l'organisation des relations entre les acteurs, enfin la recherche peut être menée sur des groupes particuliers (par localité, acteur ou domaine). Des observations exploitables par les chercheurs en sciences humaines ont ainsi pu être formulées : par exemple 80 % des liens de chaque département se dirigent vers d'autres sites du même département. La proximité physique paraît donc caractériser l'information numérique en Picardie. Néanmoins certains départements semblent plus tournés vers l'extérieur que d'autres.

La constitution d'un corpus cohérent constitue non seulement une source documentaire locale mais un reflet de ce que l'on nomme parfois « les territoires du web », notion échappant largement à la géographie physique mais qui, ramenée pour les besoins d'une enquête à un espace délimité, peut s'avérer exploitable dans une approche économique, sociologique, historique ou politique. Ainsi, le web considéré comme support et « source », à l'image de la production éditoriale archivée dans le fonds local, peut représenter un objet d'étude à part entière, révélateur des singularités et des tendances régionales.

Néanmoins, cette entreprise soulève plusieurs questions. La veille documentaire à effectuer n'est pas comparable à celle, très ciblée, qui avait été mise en place pour l'archivage du web de campagne. Si les critères de « localisation » de l'auteur ou de l'hébergeur du site d'une part, et de contenu thématique d'autre part, semblent de nature à permettre un balayage large et aisé du web, il n'en reste pas moins que l'exigence de quasi-exhaustivité requiert une veille fréquente tant ces unités documentaires sont volatiles. Même en adoptant une démarche de repérage similaire à celle de RENUPI, des

---

<sup>1</sup> Etude menée par RENUPI et l'Université de Compiègne, disponible en ligne <http://www.renupi.org> [Consulté le 12/12/2008]. Une nouvelle exploration du web picard a été lancée en 2008.

opérations manuelles sont nécessaires pour relever les sites non identifiables par le robot ; au reste il s'agissait d'une enquête statistique orientée vers des données chiffrées considérées comme globalement fiables, et non d'une démarche d'archivage. Quel qu'en soit le protocole, cette tâche sera fortement mobilisatrice pour les équipes. Bernard Huchet insiste sur le caractère « empirique » de la recherche en ligne en soulignant que « beaucoup de sites [ne sont] atteints que par un jeu complexe de ricochets et de combines » où « la patience » et « l'intuition » entrent pour une bonne part : « aucune méthodologie ne peut épuiser complètement ce phénomène. » Il ne s'agit pas, contrairement au dépôt légal de la BnF, de la capture automatisée d'un domaine identifiable.

Par ailleurs, seules les bibliothèques régionales de référence, dont le fonds local a vocation à couvrir plusieurs départements, peuvent mettre en application cette démarche avec profit. La volonté d'exhaustivité n'a guère de pertinence à l'échelle d'un simple département : elle aboutirait à une forme de balkanisation et de démultiplication des fonds numériques locaux assez étrangère à la relative harmonie issue d'une cartographie documentaire dessinée au fil du temps et somme toute respectueuse des équilibres que l'on observe dans les fonds locaux des bibliothèques et que nous évoquions dans notre deuxième partie. On peut bien entendu envisager la contribution des établissements les plus modestes à l'occasion de la constitution de corpus spécifiques, mais nous quittons alors le champ de la collecte avec volonté d'exhaustivité.

## 1.2 La sélectivité : un choix plus raisonné ?

Le choix de procéder à une sélection « raisonnée » de sites en répondant à un double objectif de continuité des collections et de qualité du contenu documentaire semble une solution adaptée aux bibliothèques plus modestes couvrant une aire géographique contenue dans les limites d'un département ou plus restreinte. La méthodologie paraît indiquée : aux divisions traditionnelles des sciences humaines on adjoindra des contenus permettant d'actualiser la documentation locale en archivant les sites institutionnels, les sites personnels, les outils collaboratifs et communautaires d'intérêt local, etc. Cependant, une recherche dans l'internet fait rapidement apparaître les limites de ces critères. Avant de les exposer, observons les champs<sup>1</sup> que l'on peut distinguer à travers l'examen d'une centaine de sites liés au Forez<sup>2</sup>.

### • Informations / Actualités

On peut repérer deux types de sites : les sites d'information générale présentant de nombreux articles sur l'actualité locale, un catalogue quasi exhaustif des manifestations culturelles, des contributions visant à mettre en valeur le patrimoine, etc. A cet égard le site *Forez Info, Portail régional Saint-Etienne Forez*<sup>3</sup> est remarquable tant par la diversité que par la qualité du contenu : il propose en outre une sélection de 1146 sites régionaux référencés et classés<sup>4</sup> orientée en grande partie vers les services proposés dans

---

<sup>1</sup> Une liste d'adresses URL est donnée à titre indicatif en annexe 3.

<sup>2</sup> Par Forez, nous entendons le département de la Loire, ainsi que les « marges » correspondant à la fois aux limites de l'ancien comté et à ce que la géographie physique rattache au Forez, à savoir la zone qui s'étend du versant oriental des Monts du Forez jusqu'aux Monts du Lyonnais. Cette aire s'étend jusqu'au nord de Roanne et empiète quelque peu sur les actuels départements du Puy-de-Dôme et de la Haute-Loire.

<sup>3</sup> Consultable à l'adresse : <http://www.forez-info.com/> [Consulté le 15/12/2008].

<sup>4</sup> A la date du 15/12/2008. Ce portail propose les divisions suivantes : Actualités, médias / Découverte, histoire et géographie / Enseignement, formation / Industrie et services professionnels / Restauration, hôtellerie, gîtes / Services publics / Commerce, artisanat / Divers [associations, cultes, politique, etc.] / Immobilier, architecture / Loisirs / Santé, médecine / Sport.

la région. Une deuxième catégorie de sites est constituée du web pratique, transposition des journaux de « petites annonces » : la qualité et l'exhaustivité de l'information y sont moins contrôlées.

- **Histoire locale et Patrimoine**

Des sites dédiés à l'histoire locale, à l'information bien documentée quoique citant assez peu ses sources, sont aisément repérables. Il faut y ajouter les sites des sociétés savantes, parfois très anciennes, dont le bulletin est archivé en bibliothèque depuis les origines : leur objectif paraît un complément à la publication papier. Le site de la Diana, société historique prestigieuse, propose des applications pédagogiques soignées de grand intérêt. Un site consacré à Saint-Chamond vaut surtout pour sa base photographique.

- **Divisions administratives**

Nous avons déjà évoqué les sites des petites municipalités qui se substituent parfois au bulletin communal et rendent compte du dynamisme local. Les sites des grandes agglomérations renferment des pages ou des sites consacrés aux institutions qu'elles financent : musées, archives municipales, etc. C'est aussi l'occasion de valoriser un patrimoine peu connu : ainsi la ville de Firminy consacre plusieurs pages à l'architecture réalisée par Le Corbusier.

- **Education / Formation**

Aux portails des universités qui exposent les formations proposées par l'établissement (dont les pôles de spécialité se définissent souvent en référence au tissu universitaire de la région) on peut adjoindre les sites des écoles, lycées et collèges qui se développent de plus en plus et servent de vitrine au travail des élèves et des enseignants. L'éventail des formations proposé dans un bassin éducatif est le témoin des besoins et des orientations d'une région.

- **Activités culturelles**

Les nombreux musées, même les plus modestes, s'exposent désormais sur l'internet. Des extraits de vidéos donnent parfois une image précieuse de l'organisation d'une exposition. Des sites thématiques sont créés afin de mettre en valeur les pôles de spécialité : ainsi le design est-il devenu l'emblème de Saint-Etienne par la volonté de la municipalité. Nous avons signalé l'intérêt des sites des théâtres que la BnF considère comme complémentaires aux archives papier. Les centres culturels sont aussi visibles sur le net. Des initiatives privées s'exposent et rendent compte de la diversité des acteurs culturels, qu'il s'agisse de troupes locales, de maisons d'éditions ou de librairies associatives.

- **Engagement politique / Web associatif**

L'expérience de collecte du web électoral en partenariat avec la BnF a montré le caractère éphémère de certains sites ou blogs créés à l'occasion d'une campagne. La plupart des liens vers les partis du département de la Loire référencés par la Bibliothèque de Lyon sont désormais inopérants. Un aspect important du web politique se développe autour du web participatif, constitué de réseaux d'expression révélateurs de l'engagement citoyen d'une population.

- **Nature, environnement**

Cette problématique « sociétale » trouve sur l'internet une tribune qui dépasse largement le cadre institutionnel ou politique. La réflexion sur le développement durable par

exemple est autant suscitée par les structures gouvernementales que par des initiatives collectives ou individuelles. Le parc naturel Livradois-Forez<sup>1</sup> ouvre le débat sur son réaménagement prochain par l'intermédiaire d'un blog.

- **Tissu économique**

Le développement des sites d'entreprise, encouragé par la Chambre de Commerce et d'Industrie, permet de broser un tableau du tissu économique local. Il va de soi que le site de Manufrance<sup>2</sup>, prolongement de la collection intégrale des catalogues conservée au fonds local, s'impose pour la Médiathèque de Saint-Etienne.

- **Aire linguistique**

De nombreux sites, dont certains présentent des contributions se rattachant à la « littérature grise »<sup>3</sup>, évoquent le francoprovençal ou « arpitan », vivant en Forez jusqu'à la fin du vingtième siècle, et le parler de la région de Saint-Etienne, le « gaga ». La constitution d'archives en ce domaine doit s'ouvrir aux sites consacrés à l'aire la plus large du francoprovençal : ainsi un site englobe Saint-Etienne dans une « arpitanie » qui s'étend jusqu'à Neuchâtel et descend jusqu'au Dauphiné en englobant une partie du Val d'Aoste et du Piémont.<sup>4</sup> Des documents « bilingues » sont proposés sur d'autres sites<sup>5</sup>, ainsi que des lexiques collaboratifs et quelques enregistrements sonores qui peuvent bien apparaître comme la version contemporaine de l'enquête phonographique lancée au début du vingtième siècle sur les routes de France par Ferdinand Brunot et dont la BnF conserve les enregistrements.

## **Ces catégories sont-elles pertinentes ?**

Que déduire de cette ébauche de classement ? Tout d'abord qu'il existe des sites dont la valeur documentaire nécessite la capture, le référencement et la valorisation auprès du public des fonds régionaux. Mais les catégories proposées sont loin de représenter la diversité des ressources du web, parce qu'elles ont été pensées en référence aux collections traditionnelles. Aussi peuvent-elles être considérées comme un outil d'enrichissement des fonds existants permettant de réaliser une première collection d'archives du web de qualité : elles se prêtent particulièrement bien à l'établissement de conventions entre les bibliothèques municipales et les auteurs en raison de leur faible volume.

Par ailleurs, ce classement témoigne d'un déplacement de certaines catégories documentaires dès lors que l'information est diffusée et présentée sur le web. Thierry Veyron<sup>6</sup> rappelle qu'un partage se fait implicitement entre les documents destinés aux Archives (tracts, *ephemera* et publications administratives) et ceux collectés par la Médiathèque (des imprimés de plus grande ampleur et hors du champ administratif). Or les sites d'administrations, nombreux et particulièrement stables, échappent à ce découpage car ce n'est plus seulement l'information officielle qui est diffusée mais c'est aussi l'institution qui se trouve elle-même mise-en-scène sur la page d'accueil, au fil des rubriques, à travers le compte-rendu des événements de sa vie interne et par le biais d'une véritable pédagogie de la politique de l'Etat et des collectivités. Dans ce domaine,

---

<sup>1</sup> <http://blog.parc-livradois-forez.org/index.php?Developpement-durable> [Consulté le 15/12/2008]

<sup>2</sup> <http://www.manufrance.eu> [Consulté le 15/12/2008]

<sup>3</sup> Ainsi le site déjà évoqué de René Merle : <http://www.rene-merle.com> [Consulté le 15/12/2008]

<sup>4</sup> <http://www.arpitania.eu> [Consulté le 15/12/2008]

<sup>5</sup> Citons par exemple [http://www.lexilogos.com/forezien\\_dictionnaire.htm](http://www.lexilogos.com/forezien_dictionnaire.htm) [Consulté le 15/12/2008]

<sup>6</sup> Conservateur responsable du Fonds régional et du Patrimoine à la Médiathèque de Saint-Etienne.

le changement de support opère une mutation du statut documentaire vers le champ du fonds régional.

Cette simple observation permet de mesurer combien le modèle de l'imprimé détourne de la réalité du web et appauvrit le champ de la recherche sur l'internet. Notre classement évite par exemple de considérer ce qu'il est convenu d'appeler la « blogosphère » comme un domaine à part entière et propose de la fractionner et de l'intégrer à toute force dans des divisions préétablies. Or cette démarche privilégie le contenu documentaire au détriment des formes d'expression et de l'interconnexion qui s'établit naturellement entre les divers « objets numériques » sur l'internet tout en postulant une unité de contenu, souvent problématique lorsqu'il s'agit d'observer un blog. C'est ignorer la singularité de nouveaux objets numériques et négliger une logique d'organisation et de diffusion que les chercheurs pourront considérer comme signifiante au même titre que le contenu documentaire.

## **2. PRENDRE EN COMPTE LE DOCUMENT WEB DANS SA SPECIFICITE**

### **2.1 Dépasser l'aporie du contenu**

Force est de constater que le bibliothécaire dont la recherche prend pour principal objet la qualité<sup>1</sup> du contenu documentaire risque de se livrer à une chasse hasardeuse. La métaphore de la vènerie, utilisée plaisamment par Bernard Huchet, nous semble révélatrice : *« aucune méthodologie ne peut épuiser complètement ce phénomène : le chasseur a besoin de connaître son gibier, d'anticiper ses comportements, de sauter d'un site à l'autre, d'affronter souvent beaucoup de bruit pour n'en tirer que de maigres bénéfices. Il y faut à la fois de la patience et de l'intuition, ce qu'on ne peut réduire à des protocoles de recherche uniformes. »*<sup>2</sup> Si cette phrase rend justice à la connaissance du « terrain » régional et à la capacité d'expertise des sélectionneurs, elle souligne également l'importance du tri qu'il est nécessaire d'effectuer et la quantité de « bruit » contre laquelle il faut lutter.

Ces remarques ne nous semblent pas anodines. En effet, bien que nombre de sites présentent une information de grande qualité, le contenu des pages du web régional est rarement en mesure de faire concurrence aux imprimés amassés au fil du temps pour le fonds régional. S'il nous est permis de risquer à notre tour une métaphore inspirée de l'univers numérique, nous pourrions dire que l'élagage auquel conduirait une politique fondée sur la richesse du contenu, dans l'arborescence du web, risquerait fort de ne retenir que quelques arbres étêtés – ou sites mutilés – d'un intérêt discutable. D'une certaine manière, la mutilation est aussi consubstantielle à la capture de sites web ; en faisant abstraction du contexte dans lequel elle s'inscrit, le sélectionneur parcellarise indûment l'information et isole des contenus, à l'image d'un bibliothécaire fantaisiste

---

<sup>1</sup> Par « qualité » nous entendons le degré de spécialisation du contenu (représenté sommairement dans notre grille par les catégories « spécialisé », « courant » et « vulgarisation »). De nombreuses grilles de sélection des sites proposent des échelles plus détaillées. Ce qui nous importe ici – au-delà de toute typologie –, c'est l'attitude qui consiste à évaluer le contenu en référence aux pratiques d'acquisition des imprimés.

<sup>2</sup> Réponse à l'enquête citée en Annexe 1.

qui s'aviserait de ne retenir que quelques numéros d'une publication périodique sans souci de continuité d'une collection.

## L'unité et l'écheveau

Or le document numérique est l'élément d'un écheveau dans lequel il ne représente qu'une figure : le développement, le prolongement ou l'écho d'une thématique partagée, en somme une unité dont il reste difficile de définir la valeur *en soi*. L'univers du web se caractérise par « *un fort potentiel de mise en relation des informations les unes avec les autres, mais aussi par l'implémentation dans les objets des processus de production de ces objets, un peu comme si le livre implémentait le processus de son édition et de sa diffusion.* »<sup>1</sup> Dans ce contexte de prolifération de l'information et de constitution d'un savoir en réseau, l'attitude qui consiste à isoler des éléments risque de paraître artificielle et assez décevante d'un point de vue documentaire. Sans compter qu'elle façonne une « mémoire du présent » peu conforme aux circuits de l'information tels qu'ils sont conçus et pratiqués par l'internaute, et tels qu'ils seront attendus par les chercheurs de demain, familiers de la structuration intellectuelle du web et de ses pratiques. A cet égard, Patrick Bazin dresse un parallèle imagé entre la bibliothèque traditionnelle et la bibliothèque électronique :

*« Lieu de rétention, la première était aussi un espace clos de représentation, où la logique interne des livres – ces capsules de texte stabilisé – était redoublée par un dispositif rigide de médiations destiné à guider le lecteur. Elle fonctionnait comme un bassin de décantation où, par des filtres successifs, relativement indépendants des contenus et conçus a priori (structure des fichiers, classement en rayons, agencement des salles), s'organisait un savoir sédimenté, hiérarchisé. Il était toujours loisible d'y plonger dans les profondeurs pour en retirer des messages exotiques, mais la poursuite de la singularité s'opérait toujours dans le cadre d'une topologie commune, d'une encyclopédie intégratrice.*

*L'espace de la bibliothèque électronique est bien différent. Il n'est pas seulement sans murs et largement ouvert à la diversité exponentielle des sources d'information accessibles en réseau. Il se déploie dynamiquement, autour des relations, de textes à textes, de termes à termes, qui naissent au fil de la lecture ou de la constitution du corpus et qui, à leur tour, produisent un méta-texte opératoire. »<sup>2</sup>*

Quelle attitude adopter afin d'intégrer à la démarche de sélection et d'archivage, activités traditionnelles de la bibliothèque fortement déterminées par le document papier, la structure dynamique du web tout en proposant de nécessaires découpages dans un gisement informationnel en voie de complexification constante ? Quelle pourrait être la place dans ce programme de documents nativement numériques tels que les hypertextes, les images de synthèse, les forums de discussion, les réseaux de sites ?

## 2.2 Problématiser et constituer des corpus

Il nous semble tout d'abord qu'une collecte du web doit éviter de créer les catégories d'un nouvel encyclopédisme. Parce que la BnF effectue des saisies régulières du domaine français dans lesquelles il sera possible de puiser et parce que la volonté

<sup>1</sup> Patrick BAZIN, « La mémoire reconfigurée », in *Cahiers de médiologie*, n°11, 2001, p. 180.

<sup>2</sup> *Idem*, p. 181-182.

d'exhaustivité en matière d'internet – même ramenée au champ régional – risque de se trouver confrontée à une prolifération difficile à anticiper, qui entraînerait des coûts de traitement et de stockage élevés.<sup>1</sup> Aussi paraît-il préférable de substituer à la question induite par la sélection des sources traditionnelles – quel contenu, intéressant pour le fonds, rechercher ? – l'interrogation suivante : quel ensemble cohérent, révélateur de thématiques régionales, puis-je prélever de manière à donner une image représentative de la structuration du web à une époque donnée ? Il s'agit en quelque sorte d'ouvrir le regard du sélectionneur à l'échelle numérique tissé autour de problématiques dont le développement sur le web présente un intérêt régional. Néanmoins, qu'on ne s'y trompe pas : la démarche consiste moins à élargir le champ à des objets plus complexes qu'à intégrer de nouvelles formes de relations, voire de sociabilités intrinsèquement liées à l'objet numérique. Comme le souligne Patrick Bazin « *l'avenir de la mémoire culturelle n'est pas uniquement une affaire de logistique. Il dépend aussi des formes de sociabilité qui se développent conjointement aux techniques. Or c'est justement sur le relationnel, sur la socialisation des pensées et des affects, que la révolution numérique a le plus d'impact. Non pas pour les atomiser davantage, comme on se plaît trop souvent à le dire, mais pour les renforcer.* »<sup>2</sup>

Ainsi, la constitution de corpus numériques doit permettre d'associer au contenu documentaire les conditions de sa production et de son développement sur le web, en restituant une image de l'aire de déploiement d'un champ thématique à travers ses dimensions sociales, militantes, etc. Bien entendu, il ne s'agit que d'une image, tant il est illusoire de penser préserver l'ensemble des hypertextes. Mais cette recombinaison visant à rester au plus près de la structuration propre à l'internet présente des virtualités d'exploitation beaucoup nombreuses qu'une sélection « atomisante » de quelques sites conforme à des principes pensés pour les supports physiques.

Quelques exemples éclaireront notre propos. Il est possible de dégager plusieurs problématiques particulièrement riches au sein de la sélection opérée dans le « web forézien » avec l'intention première de créer un choix raisonné d'objets numériques aptes à s'intégrer aux collections du fonds régional.

#### • **La préoccupation environnementale et le développement durable**

A côté des sites institutionnels (Fédération Rhône-Alpes de protection de la nature sur l'Ecopôle du Forez, Parc Naturel Livradois-Forez, Parc Naturel du Pilat, Centre ressource du Développement durable), on distingue un réseau collaboratif associant aussi bien un blog sur l'avenir du Parc Livradois-Forez qu'un « atelier citoyen de sylviculteurs » ou bien encore l'exposé au format *.pdf* de la « démarche d'établissement en développement durable » du Lycée du Forez (Feurs) qui présente la réflexion et les actions menées par la communauté éducative. On mesure à quel point cette documentation – dont le relevé n'est pas exhaustif – constitue un matériau utile tant au lecteur contemporain qu'au chercheur de demain dans la mesure où il restitue tout ensemble une démarche et son objet, les moyens d'une action (institutionnelle et individuelle), et le réseau de sociabilité créé autour d'une réflexion commune.

---

<sup>1</sup> Gildas ILLIEN insiste sur le caractère partenarial d'une collaboration entre la BnF et les bibliothèques : si les ingénieurs et techniciens doivent être « conscients des contenus qu'ils produisent », les bibliothécaires « ne peuvent quant à eux ignorer le fonctionnement des serveurs ni les coûts informatiques qui pèsent sur la production au risque de dérégler toute l'économie et la cohérence du travail. » in « Le dépôt légal de l'internet en pratique : les moissonneurs du web », *BBF* n°6, 2008, p. 24.

<sup>2</sup> Article cité, p. 183.

### • La mise en valeur du patrimoine local

Dans ce cas, il s'agirait moins de créer une collection de sites institutionnels et individuels (ces derniers faisant difficilement concurrence à la documentation papier et risquant au final de constituer un cabinet de curiosités local en marge des collections), que de mettre en relief la valorisation du patrimoine dans ses aspects institutionnels, associatifs, sociaux,... Aux pages des petits musées privés, témoins d'activités qui façonnèrent l'identité régionale, s'ajoutent de nombreuses sociétés savantes, des blogs liés au patrimoine local, des sites parfois éphémères issus de la volonté de faire connaître un aspect oublié de l'histoire régionale (ainsi la réouverture d'une ancienne ligne ferroviaire dans un département qui vit la naissance du chemin de fer français) ou bien encore des sites créés pour appeler à une mobilisation en faveur de la sauvegarde d'un site.

Des problématiques semblables pourraient être dégagées autour d'une thématique telle que la « défense et illustration de la langue francoprovençale », sans parler des perspectives qu'offrent le web militant, associatif ou politique par exemple. Peut-être nous objectera-t-on que cette attitude revient à opérer une sélection documentaire à la place des chercheurs, en délimitant *a priori* des champs d'investigation alors que la mission du bibliothécaire consiste à mettre à disposition la documentation la plus large possible. Nous pourrions répondre que la collection est une démarche consciente de sélection, et que le domaine de l'internet nous oblige à « *élargir le champ de notre mémoire à une réalité beaucoup plus composite et mouvante que celle à laquelle nous avaient habitués les musées et les bibliothèques.* »<sup>1</sup> La constitution de corpus larges, plus fidèles à la nature de l'information numérique, offre en outre trois avantages : elle est un instrument de valorisation de la bibliothèque et du fonds régional, elle peut s'inscrire aisément dans une architecture globale de l'archivage de l'internet en France, et elle offre l'opportunité d'établir de nouveaux partenariats.

## 3. PREVOIR LA RESTITUTION

### 3.1 Une offre structurée

Que l'on opte pour une relative exhaustivité dans la capture du web régional ou que l'on fasse le choix de la sélectivité (soit par l'enrichissement des catégories documentaires du fonds, soit par la constitution de corpus), l'archivage de l'internet doit se présenter comme une offre structurée. En somme, le travail du bibliothécaire consiste à faciliter l'appropriation de ces archives d'un nouveau type grâce à une triple intervention : la délimitation de champs du web d'intérêt régional, une capture affinée de certains sites afin de conserver des archives de qualité et l'accompagnement du lecteur à travers ces ressources. S'en remettre aux captures massives du domaine français assurées par la Bnf paraît insuffisant : l'archivage de certains sites d'intérêt manifeste nécessite une fréquence d'aspiration adaptée ; par ailleurs, l'importance du volume collecté appelle une valorisation active de secteurs ciblés dans le web, tant l'accumulation d'instantanés expose à la démultiplication des références et à la redondance de l'information qui font obstacle à l'exploitation de ces ressources.

---

<sup>1</sup> Patrick Bazin, *op. cit.*, p. 183.

## Quelle valorisation ?

La valorisation de ces ressources est donc essentielle, et l'on peut imaginer une configuration dans laquelle la bibliothèque municipale serait la pièce locale d'une figure beaucoup plus vaste. Des bibliothécaires envisagent un « portail inter-établissements »<sup>1</sup> qui serait l'occasion de présenter aussi bien le patrimoine numérisé que les archives du web régional. D'autres institutions partenaires pourraient y prendre place : « les archives municipales et départementales »<sup>2</sup>, les structures culturelles régionales (ainsi le « Conservatoire occitan » dans le sud-ouest), etc. Un projet de coopération entre la BnF et les Archives nationales au sujet du web administratif pourrait être l'axe central d'une redistribution régionale de cette catégorie d'archives à travers les points d'accès qui s'ouvriront probablement à terme dans les BDLI. Un premier schéma d'archivage et de signalement en bibliothèque municipale pourrait s'appuyer sur deux ensembles documentaires distincts : le référencement de sites relativement stables et essentiels à la documentation régionale (musées, institutions, etc.) et les corpus régionaux, révélateurs de problématiques locales, d'autre part.

## Signalement

Si un référencement permettant d'intégrer ces objets numériques aux ressources traditionnelles du catalogue est légitime, il ne faut pas oublier que le décret d'application de la loi DADVSI ne pourra envisager qu'une extension des points de consultation sans possibilité de cession ou de duplication des archives numériques. Par conséquent, seules les archives acquises par convention avec leur « créateur » pourront bénéficier de ce traitement. Il est néanmoins souhaitable que le signalement des archives du web, assorti d'une brève description qui pourra s'appuyer sur les indications de la grille de sélection, soit associé aux dossiers documentaires constitués par le fonds régional de la bibliothèque. En outre, l'accès aux collections de la BnF permet une consultation de l'archive qui recrée son univers numérique d'origine :

*« Le signalement des données répond lui aussi à une logique de masse. Le catalogage des sites archivés a été exclu au profit de l'indexation automatique. Toutes les collections de la BnF sont aujourd'hui indexées pour une recherche par URL avec le logiciel Wayback Machine : on peut retrouver un site si l'on connaît son adresse exacte. [...] Et c'est cette indexation qui permet de naviguer dans les archives comme sur le web vivant, en cliquant de lien en lien. »<sup>3</sup>*

Ce contexte juridiquement contraignant crée paradoxalement pour les bibliothèques municipales dépositaires les conditions d'une restitution optimale des archives de l'internet. Dans cette configuration qui vise à reconstituer l'environnement original des documents numériques, il est intéressant d'observer que le rôle du bibliothécaire tend à se déplacer du traitement des données vers la fonction « éditoriale »<sup>4</sup> qui consiste à valoriser des corpus, à suggérer des parcours et à « mettre en scène »<sup>5</sup> la documentation collectée. Les volumes pris en compte ne permettent plus guère le référencement exhaustif et détaillé d'unités documentaires dont la délimitation s'avère souvent problématique.

---

<sup>1</sup> Réponse à l'enquête citée en Annexe 1.

<sup>2</sup> Une autre réponse à la même enquête.

<sup>3</sup> Gildas ILLIEN, article cité, p. 23.

<sup>4</sup> L'expression est de Gildas ILLIEN, article cité, p. 23.

<sup>5</sup> *Idem*, p. 23.

De surcroît, la constitution de corpus thématiques, qui donne lisibilité et cohérence aux archives du web, peut être l'occasion d'un contact renouvelé avec les publics spécifiques du fonds local.

### **3.2 La bibliothèque municipale prestataire ?**

L'élaboration de corpus thématiques peut trouver son origine dans la connaissance des réalités régionales et dans une expertise du web menée par les bibliothécaires. Elle peut aussi être considérée comme une anticipation des besoins futurs des chercheurs autour de problématiques historiques, sociales, économiques. Par ce biais, les regroupements documentaires sont susceptibles de renouveler l'intérêt des lecteurs. Une voie féconde consisterait certainement à prendre appui sur les partenariats développés par nombre de bibliothèques avec les publics spécifiques du fonds régional : les scolaires, les étudiants et les chercheurs.

#### **Tisser des partenariats**

Une étude récente menée à la Bibliothèque municipale de Lyon sur les publics des fonds locaux et régionaux soulignait notamment combien il importe de « reconquérir le public universitaire »<sup>1</sup> grâce à une communication accrue autour des collections qui fasse « connaître les gisements documentaires inexploités » en privilégiant des approches multiples. De nombreux établissements organisent déjà des visites à l'intention des scolaires et des étudiants avancés auxquels sont parfois signalés des champs de recherche. La bibliothèque est souvent le lieu de rencontres ou de colloques réunissant enseignants-chercheurs, étudiants et plus large public. Néanmoins, il reste à retisser un lien privilégié avec une Université souvent perçue comme trop « indifférente »<sup>2</sup>. Les corpus régionaux du web sont une opportunité de développer une offre en ce sens ou bien de répondre aux besoins de chercheurs et de scolaires en faisant à l'occasion de la bibliothèque le prestataire de services « à la demande ».

On perçoit sans peine l'intérêt que peut présenter une documentation régionale cohérente portant sur des problématiques économiques ou sociales pour le public lycéen des sections de SES<sup>3</sup>, ou bien les étudiants des sections de techniciens supérieurs dont une partie de la formation repose sur la sensibilisation à l'environnement régional et sur des études de cas de manière à anticiper l'insertion dans la vie de l'entreprise. Par ailleurs, les orientations actuelles de l'enseignement secondaire privilégient l'éducation citoyenne et la transversalité : pour ne reprendre que les deux exemples cités – « la préoccupation environnementale et le développement durable », « la mise en valeur du patrimoine local » –, ces corpus constituent à l'évidence un champ d'investigation de nature à susciter l'intérêt des enseignants et des élèves. En outre, des disciplines universitaires telles que les sciences politiques, la sociologie, la linguistique peuvent trouver dans la documentation régionale la matière de recherches à mener. Dans cette perspective, des partenariats formalisés sont à envisager entre la bibliothèque et des équipes universitaires prescriptrices, à condition que le bénéfice de la sélection des ressources effectuée par les bibliothécaires intéresse le fonds local. C'est là un moyen

---

<sup>1</sup> Laure RIOUST, *op. cit.*, p. 66.

<sup>2</sup> *Idem*, p. 66.

<sup>3</sup> Les SES (Sciences économiques et sociales), mêlant analyse économique et initiation à la sociologie, sont, dans la maquette actuelle des enseignements du lycée, une option en classe de seconde et une matière de spécialité pour une filière du baccalauréat en classe de première et en classe terminale.

d'enrichissement des collections autour de problématiques signalées par les chercheurs et l'on peut à l'occasion solliciter une aide financière des unités de recherche si la sélection nécessite une mobilisation prolongée du personnel.

Le développement et le renforcement de la collaboration entre les bibliothèques et les structures d'enseignement et de recherche sous la forme de partenariats menés à la faveur de projets à thématique régionale constitueraient un gain incontestable en termes de valorisation de la bibliothèque et de sensibilisation des publics aux ressources du fonds local.

## **Organiser l'accompagnement de l'utilisateur**

La prestation fournie par la bibliothèque ne peut cependant pas se limiter à la mise à disposition de corpus cohérents. Si l'environnement recréé par la BnF permet une navigation sur le web archivé aussi proche que possible des conditions habituelles de l'utilisation de l'internet, il serait pourtant imprudent de s'en remettre à la pratique empirique de l'internaute. Parce que certaines archives sont lacunaires et que le déficit d'information doit être compensé dans ce cas par une modification de la stratégie de recherche : aux réflexes de la requête intuitive menée sur le web vivant méritent d'être associées des procédures plus construites qui tiennent compte de la nature figée du document d'archives. Appréhender une archive du web revient tout d'abord à consulter une liste de saisies d'un même site s'étalant sur une période très variable selon la pérennité du document. Les mises à jour sont habituellement signalées sur la *Wayback Machine* à l'aide d'un astérisque. Il convient donc à chaque utilisateur d'opérer une première sélection des versions archivées en fonction de sa recherche et de la nature du site qu'il consulte, dont il peut observer la fréquence de renouvellement.

Il ne faut pas oublier que le public familiarisé très jeune avec les nouvelles technologies de l'information a développé un usage intuitif et personnel lui permettant d'accéder rapidement au contenu et que les archives de l'internet peuvent constituer un obstacle à la recherche en raison d'une redondance de l'information qui par certains aspects apparente ces ressources aux archives traditionnelles dont le défrichage exige méthode et patience. Il est donc souhaitable que le bibliothécaire expose brièvement la spécificité des archives de l'internet et les difficultés qu'elles peuvent poser de manière à proposer quelques procédures qui évitent à l'utilisateur le découragement et les déconvenues. Un accompagnement personnalisé, d'ailleurs conçu comme le prolongement des services généralement offerts par le fonds régional à ses lecteurs, doit prendre le relais de cette présentation méthodologique et soutenir l'utilisateur dans le traitement de ces documents d'archives.



## Conclusion

---

Dans l'esprit de nombreux bibliothécaires les archives du web ont désormais acquis une légitimité suffisante pour figurer au côté des ressources traditionnelles du fonds local dont elles constituent en partie le prolongement. Cependant la loi de 2006, en même temps qu'elle offre un cadre juridique au dépôt légal du web, contraint sévèrement les conditions de la collecte et de la restitution. Pour l'heure, seuls la BnF et l'INA sont autorisés à effectuer des captures dans l'internet sans solliciter l'assentiment des ayants droit. Aussi quelques établissements s'essayaient à rédiger des conventions – avec des créateurs de sites institutionnels dans un premier temps – mais l'on reste encore loin d'une démarche de construction d'une collection.

Si la voie reste étroite pour les bibliothèques municipales, notre conviction est qu'il y a place pour un schéma national d'archivage de l'internet prenant appui sur les connaissances des bibliothécaires en région et bénéficiant de l'infrastructure technique de la BnF qui permet d'assurer l'aspiration des sites, leur indexation, leur mise à disposition et leur conservation, autant de tâches que les établissements municipaux ne sont plus en mesure d'assurer en raison des volumes traités dès lors qu'il s'agit de l'internet, sans compter l'importance des espaces de stockage nécessaires, la technicité des opérations et les moyens de garantir la pérennité des archives dans un contexte de rapide évolution des supports. Les campagnes d'archivage du web électoral ont déjà permis d'expérimenter une collaboration fructueuse entre la BnF et les BDLI, établissements qui seront probablement dotés, à terme, d'un point de consultation des archives de l'internet. On peut ainsi envisager un réseau d'archivage régional fondé sur l'échelon intermédiaire que représente la BDLI, établissement en mesure de fédérer les demandes des bibliothèques de la région et d'assurer l'interface avec la BnF. Néanmoins la restitution des ressources capturées demeure incertaine en bibliothèque municipale et il est souhaitable d'associer à l'effort fourni le soutien des agences de coopération dont les missions s'accordent avec l'objectif de valorisation du patrimoine numérique régional.

Cette architecture est destinée à faciliter la coopération à l'échelle nationale en matière d'archivage du web : elle permet de s'adapter aux contraintes juridiques mais elle ne peut se substituer à une réflexion méthodologique sur les finalités d'un tel projet. Il s'agit d'une part – côté bibliothèques municipales – de référencer le web régional pour l'intégrer aux collections du fonds local, et d'autre part – côté BnF – de donner forme et visibilité aux ressources numériques conservées, issues de captures larges et difficilement valorisables en l'état. Deux options sont envisagées par les bibliothécaires en région : la recherche d'une certaine exhaustivité qui apparente la collecte à un « dépôt légal régional » du web conçu comme la continuité des acquisitions d'une bibliothèque à forte vocation régionale, ou bien la sélectivité, choix raisonné de ressources se distinguant par la qualité de leur contenu ou par le caractère spécifique des problématiques locales qu'elles abordent.

Un traitement semi-automatisé convient au premier type de collecte. Néanmoins un suivi régulier et permanent s'impose de manière à assurer la constitution d'une collection cohérente parce qu'il est nécessaire d'adapter la fréquence de capture de certains sites et parce que la recherche humaine doit pallier les limites du robot. Cette démarche peut ouvrir la voie à une redistribution des ressources du web local – en associant les Archives pour le web administratif – sur des portails régionaux. Elle s'adresse avant tout aux BMVR dont l'exhaustivité en matière de documentation locale est affirmée, et aux BDLI au sein desquelles il sera possible à terme de consulter ces archives.

La sélectivité semble donc réservée aux bibliothèques municipales. Il ne s'agit pourtant pas d'une option *a minima*. Si l'on écarte une conception trop étroite qui se limiterait à considérer les documents de l'internet comme une simple transposition des collections physiques sur un autre support, il est possible de délimiter des corpus qui mettent en relief les caractéristiques d'un territoire et sa dynamique particulière. Encore faut-il prendre soin de préserver la cohérence d'ensembles qui possèdent leur logique propre sur l'internet. Cette démarche peut donner lieu à des collectes ponctuelles aussi bien qu'à des recherches sur le long terme mais ciblées ; par conséquent elle s'avère moins mobilisatrice pour les personnels. De surcroît, elle est de nature à s'insérer sans difficulté dans un réseau régional d'archivage et elle se prête tout particulièrement à des campagnes nationales. Enfin, ces corpus thématiques paraissent aisément valorisables auprès des publics des fonds locaux, notamment lorsqu'ils sont constitués en réponse à des demandes spécifiques.

La réflexion sur l'archivage du web et sur la mémoire numérique permet d'appréhender certaines évolutions professionnelles liées aux nouveaux modes de production et de diffusion de l'information. Face aux volumes pris en compte, et dans un contexte de croissance exponentielle des ressources, la mise en place d'une architecture nationale et le recours à une solide infrastructure technique semblent impératifs. Par ailleurs, les procédures automatisées doivent prendre le relais du traitement de l'unité documentaire que l'on réservera à des corpus limités. Du reste les contenus électroniques sont déjà configurés dans une perspective d'accès et de traitement. Pour autant, les compétences professionnelles des bibliothécaires ne s'en trouvent pas reléguées au second plan. Bien au contraire, cette évolution nécessite un recentrement sur la capacité d'expertise des bibliothécaires et sur leur ancrage local afin de définir les problématiques régionales qui méritent une recherche. Si la conservation du numérique se déroule désormais hors du champ de la bibliothèque municipale, cette dernière doit se penser de plus en plus comme prestataire de services auprès de ses publics et il importe désormais d'ajouter à la valorisation active des ressources une réflexion sur l'internet considéré comme objet d'étude à part entière, en tant que vecteur de l'information et diffuseur de modes d'expression originaux qui donnent naissance à des réseaux de sociabilité dont on ne retrouve pas l'équivalent hors de l'internet.

C'est donc à un double impératif, paradoxal en apparence, que semble renvoyer l'archivage du web en bibliothèque municipale. Habitée à se penser comme le centre d'un territoire, la bibliothèque doit simultanément faire l'effort d'un décentrement volontaire pour se projeter dans la logique d'un réseau de collecte plus vaste au sein duquel sa recherche singulière prendra sens. N'est-ce pas en somme l'image spéculaire de l'archive électronique elle-même dont le contenu n'offre qu'un intérêt limité s'il n'est restitué dans son contexte originel ? Aussi peut-on affirmer avec Louise Merzeau qu'«archiver le web, c'est reconnaître que le capital symbolique à transmettre ne

*saurait désormais se couper des flux de données sans perdre une part essentielle de sa dynamique et de ses contenus. Mais c'est aussi exposer l'archive à une contamination par l'éphémère, qui tend à modifier la logique et la portée du geste patrimonial.<sup>1</sup>»*

---

<sup>1</sup> Louise MERZEAU, in « Web en stock » *Cahiers de médiologie*, 2003, n° 16.



# Bibliographie

---

## ARCHIVER LE WEB : PROJETS ET METHODES

**Association Française pour le Nommage Internet en Coopération (AFNIC).** Statistiques des domaines en .fr et .re [en ligne]. [Consulté le 26/10/2008].

URL : <<http://www.afnic.fr/actu/stats>>

**ASCHENBRENNER, Andreas.** *Long-Term Preservation of Digital Material : Building an Archive to Preserve Digital Cultural Heritage from the Internet* [en ligne]. Thèse de doctorat.[Consulté le 10/12/2008].

URL : <<http://www.ifs.tuwien.ac.at/~aola/publications/thesis-ando/>>

**BASSET, Hervé.** *Sélection et évaluation de sites web scientifiques* [en ligne]. Mémoire de maîtrise, Université catholique de l'Ouest, Angers, 2000. [Consulté le 26/10/2008]

URL : <[http://australie.uco.fr/info/biblio-info/menu1/menu1\\_2/memoire\\_herve.pdf](http://australie.uco.fr/info/biblio-info/menu1/menu1_2/memoire_herve.pdf)>

**BAZIN, Patrick.** La mémoire reconfigurée. *Cahiers de médiologie*, 2001, n°11, p. 177-183. [en ligne] [consulté le 07/09/2008]

URL : <[http://www.mediologie.org/collection/11\\_transmettre/bazin.pdf](http://www.mediologie.org/collection/11_transmettre/bazin.pdf)>

**BRYGJFELD, Svein Arne, BARYLA, Christiane (trad.).** Accès aux archives du web : le projet d'accès aux archives nordiques du web [en ligne]. *68<sup>th</sup> IFLA Council and General Conference*, Galsgow : 2002. [Consulté le 07/09/2008]

URL : <<http://www.ifla.org/IV/ifla68/papers/090-163f.pdf>>

**CARDON, Dominique.** Fiche de présentation «Membre associé – chercheur au laboratoire des usages de France Telecom R&D » [en ligne]. Paris : Ecole des Hautes Etudes en Sciences Sociales. [Consulté le 10/12/2008].

URL : <<http://cems.ehess.fr/document.php?id=155>>

**CHAUSSARD, Audrey, GHARSALLAH, Mehdi, MONTFORT, Jérôme.** *Panorama mondial de l'archivage du web* [en ligne]. Serda, pour le Ministère de la Région Wallonne. Bruxelles : Serda, 2003. [Consulté le 10/12/2008]

URL :

<[http://easi.wallonie.be/servlet/Repository/panorama\\_mondial\\_du\\_web.pdf?IDR=8605](http://easi.wallonie.be/servlet/Repository/panorama_mondial_du_web.pdf?IDR=8605)>

**ERTZSCHEID, Olivier.** *Question de taille...*[en ligne]. In Affordance.info. [Consulté le 07/12/2008]

URL : <[http://affordance.typepad.com/mon\\_weblog/2007/03/question\\_de\\_tai.html](http://affordance.typepad.com/mon_weblog/2007/03/question_de_tai.html)>

**GUICHARD, Eric.** *Equipe Réseaux, Savoirs et Territoires. Bilans et projets quadriennaux (2002-2005 et 2006-2009)* [en ligne]. Ecole Normale Supérieure. [Consulté le 10/12/2008].

URL : <<http://barthes.ens.fr/equipe/QuadriERST2005.pdf>>

**HAETTIGER, Magali.** *L'Archivage des sites web d'intérêt régional.* Mémoire d'étude, Villeurbanne : ENSSIB, 2003. [en ligne] [Consulté le 10/07/2008]

URL : <<http://enssibal.enssib.fr/bibliotheque/documents/dcb/haettiger.pdf>>

**HOOG, Emmanuel.** Internet a-t-il une mémoire ? *Le Monde*, 17 août 2002.

**Internet Archive** (Association à but non lucratif, San Francisco, Etats-Unis) *Wayback Machine* [en ligne]. [Consulté le 10/12/2008]

URL : <http://www.archive.org/web/web.php>

**Internet Archive** (Association à but non lucratif, San Francisco, Etats-Unis) *Robot crawler Heritrix* [en ligne]. [Consulté le 10/12/2008]

URL : <<http://crawler.archive.org/>>

**KAVCIC-COLIC, Alenka.** Archiving the Web : Some legal Aspects [en ligne]. 68<sup>th</sup> IFLA Council and General Conference, Glasgow. [Consulté le 07/09/2008]

URL : <<http://www.ifla.org/IV/ifla68/papers/116-163e.pdf>>

**Library of Congress.** Web capture [en ligne]. [Consulté le 10/12/2008]

URL : <<http://www.loc.gov/webcapture/>>

**LUPOVICI, Christian.** The international Internet Preservation Consortium : General strategy and state of work [en ligne]. *Conference of European National Librarians, Helsinki*, 27 septembre 2007. [Consulté le 18/10/2008]

URL : <<http://netpreserve.org/events/IIPC%20CENL%20070927.ppt>>

**LYMAN, Peter.** Archiving the world wide web [en ligne]. *Council on library and information resources. Building a National Strategy for Preservation : Issues in Digital Media Archiving.* Berkeley, 2002. [Consulté le 10/09/2008]

URL : <http://www.clir.org/pubs/reports/pub106/web.html>

**Médiamétrie.** *Mesure d'audience et étude des médias audiovisuels et interactifs.* [en ligne] [Consulté le 10/12/2008]

URL : <<http://www.mediametrie.fr/>>

**MERZEAU, Louise.** Web en stock [en ligne]. *Cahiers de médiologie*, 2003, n° 16. [Consulté le 10/09/2008]

URL : <<http://www.merzeau.net/txt/memoire/webenstock.html>>

**Organisation des Nations-Unies pour l'Éducation, la Science et la Culture (UNESCO).** *Charte sur la conservation du patrimoine numérique* [en ligne]. Paris : UNESCO, 2003. [Consulté le 10/09/2008]

URL : <<http://portal.unesco.org/ci/en/ev.php->

[URL\\_ID=13367&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=13367&URL_DO=DO_TOPIC&URL_SECTION=201.html)>

**Raynette.** Moteur WHOIS de recherche d'un domaine existant (propriétaire, informations DNS/IP, serveur) [en ligne] [Consulté le 10/12/2008].

URL : <<http://www.raynette.fr/services/whois/>>

**RUSSELL, Kelly.** *Why can't We Preserve Everything ? Selection Issues for the Preservation of Digital Materials* [en ligne]. St Pancras : The British Library, 1999.

[Consulté le 10/09/2008]

URL : <http://www.leeds.ac.uk/cedars/documents/ABS01.htm>

**UK Web Archiving Consortium.** *Project overview* [en ligne]. [Consulté le 10/12/2008]

URL : <<http://info.webarchive.org.uk/>>

**Université de Compiègne.** *RENUPI, Repères numériques de Picardie* [en ligne]. Observatoire des technologies de l'information sur la région Picardie. [Consulté le 10/12/2008].

URL : <<http://www.renupi.org/>>

**Université de Nantes.** *Outils et techniques pour explorer le web invisible* [en ligne]. Nantes : Laboratoire IRIN. [Consulté le 10/12/2008].

URL : <[http://www.sciences.univ-](http://www.sciences.univ-nantes.fr/info/perso/permanents/desmontils/InvisibleWeb.html)

[nantes.fr/info/perso/permanents/desmontils/InvisibleWeb.html](http://www.sciences.univ-nantes.fr/info/perso/permanents/desmontils/InvisibleWeb.html)>

**VERHEUL, Ingeborg.** *Networking for Digital Preservation : Current Practices in 15 National Libraries* [en ligne]. Koninklijke Bibliotheek / IFLA. Munich : K. G. Saur, 2006. [Consulté le 10/12/2008]

URL : <http://www.ifla.org/VI/7/pub/IFLAPublication-No119.pdf>

**XitiMonitor.** *Etude du comportement de l'internaute. Régions françaises* [en ligne]. Société AT Internet, Bordeaux Mérignac. [Consulté le 10/12/2008]

URL : <<http://www.xitimonitor.com/pdf/fr-FR/Régions%20françaises%20-%20Mai%202007.pdf>>

## LE CADRE JURIDIQUE EN FRANCE

**Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information** [en ligne] [Consulté le 10/09/2008]

URL : <[http://admi.net/eur/loi/leg\\_euro/fr\\_301L0029.html](http://admi.net/eur/loi/leg_euro/fr_301L0029.html)>

**Loi n° 2006-961 du 1<sup>er</sup> août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information** [en ligne] [Consulté le 10/09/2008]

URL :

<<http://www.legifrance.gouv.fr/WAspad/UnTexteDeJorf?numjo=MCCX0300082L>>

**ALIX, Yves.** Les exceptions bénéficiant aux bibliothèques, la révolution ? *Revue Lamy Droit de l'immatériel*, mars 2007, n° 25, pp. 25-28.

**BENABOU, Valérie-Laure.** Patatras ! A propos de la décision du Conseil constitutionnel du 27 juillet 2006, *Propriétés intellectuelles*, juillet 2006, n° 20, p. 240-242.

**GEIGER, Christophe.** La loi du 1<sup>er</sup> août 2006, une adaptation du droit d'auteur aux besoins de la société de l'information ? *Revue Lamy Droit de l'immatériel*, mars 2007, n°25, pp. 67-77.

**GLEIZE, Bérangère.** La culture à l'épreuve de la loi du 1<sup>er</sup> août 2006 (et vice-versa). In Bruguère, Jean-Michel (dir.), *Droit d'auteur et culture*, Paris : Dalloz, 2007.

**LAHARY, Dominique,** Les bibliothèques et la loi DADVSI : survivre dans un débat fracassant. *Bulletin des bibliothèques de France*, 2006, t. 51, n°5.

**MAUREL, Lionel,** *Bibliothèques numériques : le défi du droit d'auteur*. Villeurbanne : Presses de l'ENSSIB, 2008.

**VALETTE, Arnaud.** *L'impact de la loi DADVSI sur l'exploitation des contenus numériques* [en ligne]. Villeurbanne, ENSSIB, « Bibliothèque numérique », 2006. [Consulté le 10/09/2008]

URL : <[www.enssib.fr/bibliotheque-numerique/document-1155](http://www.enssib.fr/bibliotheque-numerique/document-1155)>

**VIVANT, Michel,** Les exceptions nouvelles au lendemain de la loi DADVSI. *Recueil Dalloz*, 14/09/2006, n° 31, pp. 2159-2163.

## **LA BNF ET LE DEPOT LEGAL DU WEB**

**Bibliothèque nationale de France (BnF).** *Dossier de presse. Internet en campagne : de l'archivage des sites électoraux à leur communication au public.* [en ligne] Paris : BnF, 2006. [consulté le 26/10/2008]

URL : <[http://www.bnf.fr/pages/presse/dossiers/internet\\_campagne.pdf](http://www.bnf.fr/pages/presse/dossiers/internet_campagne.pdf)>

**Bibliothèque nationale de France (BnF), Institut National de l'Audiovisuel (INA).** *Dossier de presse. Les enjeux du dépôt légal de la Toile* [en ligne] Paris : BnF, INA, 2005. [consulté le 26/10/2008]

URL : <<http://www.bnf.fr/pages/presse/dossiers/toile.pdf>>

**Bibliothèque nationale de France (BnF).** *Elections 2007 : réunion des sélectionneurs des BDLI – compte-rendu.* Paris : BnF, 19 octobre 2007. (Compte-rendu de réunion. Réf. BnF-ADM-2007-011019-01)

**GHARSALLAH, Mehdi.** *Archivage du web français et dépôt légal des publications électroniques* [en ligne]. Documentaliste, Sciences de l'Information, 2004. [Consulté le 10/09/2008]

URL : <[http://archivesic.ccsd.cnrs.fr/sic\\_00001311.html](http://archivesic.ccsd.cnrs.fr/sic_00001311.html)>

**ILLIEN, Gildas, GAME, Valérie,** Le dépôt légal d'Internet à la Bibliothèque nationale de France : Cadre juridique, modèle de collecte, évolutions des métiers, *Bulletin des bibliothèques de France*, 2006, n° 3.

**ILLIEN, Gildas,** Le dépôt légal de l'internet en pratique : les moissonneurs du web, *Bulletin des bibliothèques de France*, 2008, n° 6.

**ILLIEN, Gildas**, Repenser la politique documentaire à l'échelle de la Toile. : l'expérience du dépôt légal d'Internet à la BnF [en ligne]. *Congrès de l'ADBU, septembre 2007*. [Consulté le 19/11/2008]  
URL : <[http://www.adbu.fr/IMG/pdf/Gildas\\_1.pdf](http://www.adbu.fr/IMG/pdf/Gildas_1.pdf)>

**LAGOUTE, Catherine, OURY, Clément**. *Elections 2007 : réunion des sélectionneurs des BDLI (30/03/2007) – manuel utilisateur de l'outil*. Paris : Bibliothèque nationale de France, Direction des Services et des réseaux, département de la bibliothèque numérique (Note. Réf. BnF-ADM-2007-010005-01)

**LUPOVICI, Catherine**. La mise en œuvre du dépôt légal électronique [en ligne]. *Journées internet pour le droit*, Paris : 3-5 novembre 2004. [Consulté le 10/09/2008]  
URL : <<http://www.frlii.org/spip.php?article48>>

**LUPOVICI, Catherine**. La conservation des publications électroniques et du dépôt légal [en ligne]. *Information for All Programme (IFAP)*. UNESCO, 2007. [Consulté le 10/09/2008]  
URL:<[http://portal.unesco.org/ci/en/ev.php-URL\\_ID=24441&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=24441&URL_DO=DO_TOPIC&URL_SECTION=201.html)>

**OURY, Clément**. *Elections 2007 : Bilan de l'organisation du projet*. Paris : Bibliothèque nationale de France, Direction des Services et des réseaux, département de la bibliothèque numérique, 28 septembre 2007. (Réf. BnF-ADM-2007-031419-01)

**OURY, Clément**. Archiver les sites de l'Internet électoral de 2007 : une illustration à l'échelle nationale du dépôt légal de la Toile, *11<sup>e</sup> journées des Pôles associés*, 22-23 novembre 2007, Strasbourg (document BnF non référencé).

**OURY, Clément, LAGOUTE, Catherine**. *Elections 2007 : Bilan documentaire du projet*. Paris : Bibliothèque nationale de France, Direction des Services et des réseaux, département de la bibliothèque numérique, 28 septembre 2007. (Réf. BnF-ADM-2007-034655-01)

**STASSE, François**. *Rapport au Ministre de la Culture et de la communication sur l'accès aux œuvres numériques conservées par les bibliothèques publiques*, avril 2005 [en ligne]  
URL :<<http://www.culture.gouv.fr/culture/actualites/rapports/stasse/stasse.rtf>>  
[Consulté le 26/10/2008]

## **POLITIQUE DOCUMENTAIRE ET FONDS REGIONAUX**

**ARNOULT, Jean-Marie**. Conservation et valorisation, ou du bon usage de la mémoire bien conservée. *Bulletin d'information de l'Association des bibliothécaires français*, 3<sup>ème</sup> trimestre 1993, n° 160.

**BERTHOMIER, Françoise**. *Les documents d'intérêt régional en bibliothèque de lecture publique : l'exemple de la bibliothèque municipale de Lyon*. Mémoire d'étude, Villeurbanne : ENSSIB, 2006.

**BLONDIAUX, Emmanuelle.** *Le fonds local et régional entre mythes et politiques : l'exemple de la salle Bretagne de la médiathèque de Lorient.* Mémoire d'étude, Villeurbanne : ENSSIB, 1996.

**BONTEMPS, Isabelle.** *Quelle politique documentaire pour l'acquisition de liens Internet en bibliothèque ?*, Mémoire d'étude, Villeurbanne : ENSSIB, 1999.

**BROCHARD, Jean-Christophe.** *Après les inaugurations, quel avenir pour les missions régionales des BMVR ?* Mémoire d'étude, ENSSIB, Villeurbanne, 2005.

**CALENGE, Bertrand (dir.).** *Bibliothécaire, quel métier ?* Paris : Editions du Cercle de la Librairie, 2004.

**CALENGE, Bertrand.** *Bibliothèques et politiques documentaires à l'heure d'Internet.* Paris : Editions du Cercle de la Librairie, 2008

**CALENGE, Bertrand.** *Conduire une politique documentaire.* Paris : Editions du Cercle de la Librairie, 1999.

**CALENGE, Bertrand.** *Accueillir, orienter, informer : l'organisation des services aux publics dans les bibliothèques.* Paris : Editions du Cercle de la Librairie, 1996.

**CALENGE, Bertrand.** *Les Politiques d'acquisition.* Paris : Editions du Cercle de la Librairie, 1997

**Centre régional du Livre-Centre et l'ABF-Centre.** *Les fonds locaux et régionaux. Synthèse des interventions.* Journée d'étude du 5 novembre 1998, Châtelleraut : non-publié.

**DEMOTIER-PILLET, Sylvie.** *Actualisation de la politique documentaire de la documentation régionale de la bibliothèque municipale de Lyon. Ouverture sur les ressources extérieures (sites Internet) : opportunités, critères, méthode.* Mémoire d'étude, Villeurbanne : ENSSIB, 2002.

**FEREY-PFALZGRAF, Nadine.** *Acquisition et collecte des documents au service de documentation régionale de la bibliothèque municipale de Lyon.* Mémoire d'étude, Villeurbanne : ENSSIB, 1993.

**GALAUD, Nicolas.** *Patrimoine écrit et identité régionale. Bulletin du Centre régional du Livre de Champagne-Ardenne, 1<sup>er</sup> semestre 1996, corps 12.*

**GUINARD, Pierre.** *Pratiques patrimoniales de la bibliothèque municipale de Lyon, Bulletin des bibliothèques de France, 1996, n°3*

**HAUCHECORNE, François.** *Fonds local et régional. Bulletin des bibliothèques de France, 1982, t. 27, n° 1.*

**IBAR, Marie-Agnès.** *Constituer un fonds local : l'exemple de la Bibliothèque municipale de Versailles.* Mémoire d'étude, Villeurbanne : ENSSIB, 1999

**JOCTEUR-MONTROZIER, Yves.** « Les fonds locaux » in *Le Patrimoine. Histoire, pratiques et perspectives*, dir. Jean-Paul Oddos. Paris : Editions du Cercle de la Librairie, 1997

**LE BERRE, Yves.** Patrimoine écrit et identité régionale. *Enrichir le patrimoine des bibliothèques en région*. Actes du colloque national organisé à Rennes, les 30 novembre et 1<sup>er</sup> décembre 1996 par l'Agence de coopération des bibliothèques de Bretagne, Rennes : éditions Apogée, 1996.

**MARTIN, Henri-Jean.** Une vision totale du livre. *Bulletin des bibliothèques de France*, 2004, n° 5.

**MOUREN, Raphaële (dir.).** *Manuel du patrimoine en bibliothèque*. Paris : Editions du Cercle de la Librairie, 2007.

**ODDOS, Jean-Paul (dir.).** *Le Patrimoine. Histoire, pratiques et perspectives*. Paris : Editions du Cercle de la Librairie, 1997.

**POULLIAS, Marie-Thérèse.** « Les logiques d'acquisition d'un grand établissement. » *Enrichir le patrimoine des bibliothèques en région*. Actes du colloque national organisé à Rennes, les 30 novembre et 1<sup>er</sup> décembre 1996 par l'Agence de coopération des bibliothèques de Bretagne, Rennes : éditions Apogée, 1996.

**RIOUST, Laure.** *De l'érudit aux usagers. Publics des fonds locaux et régionaux en bibliothèque municipale : évolutions, pratiques et représentations*. Mémoire d'étude, Villeurbanne : ENSSIB, 2008.

**RONSIN, Albert, SIMON, Nicole, WEBER, Yvette, PERETTI, Pierre de.** « La bibliothèque, mémoire de la vie locale d'hier et d'aujourd'hui. » *Lecture et bibliothèques publiques*. Actes du colloque d'Hénin-Beaumont, 20-21 novembre 1981, Lille : Office régional de la culture et de l'éducation permanente, 1982.

**RONSIN, Albert.** La bibliothèque, mémoire de la vie locale d'hier et d'aujourd'hui. *Médiathèques publiques*, 1984, n° 69.

**SYREN, André-Pierre.** Les fonds locaux de Champagnes-Ardennes. *Bulletin du Centre Régional du Livre*, 1993.

**SYREN, André-Pierre.** La bibliothèque localement universelle. *Bulletin des bibliothèques de France*, 1995, t. 40, n° 3.

**WEBER, Yvette.** Les collections régionales à la Bibliothèque Municipale de Lyon. *Bulletin des bibliothèques de France*, 1997, n° 2,



## **Table des annexes**

---

<b>ANNEXE 1.....</b>	<b>77</b>
<b>ANNEXE 2.....</b>	<b>79</b>
<b>ANNEXE 3.....</b>	<b>81</b>



# **Annexe 1**

## **QUESTIONNAIRE ENVOYE AUX BIBLIOTHEQUES MUNICIPALES EN AOUT 2008**

*Aux Bibliothèques titulaires du dépôt légal imprimeur (BDLI) ayant participé à la campagne 2007 de collecte des sites électoraux organisée par la BnF :*

1/ Lors de la collecte des sites électoraux en 2006 et 2007, quel a été le rôle exact de votre établissement ?

*Aux Bibliothèques municipales (y compris les BDLI à partir de la question 2)*

1/ Votre établissement a-t-il envisagé une collecte de sites régionaux, sur le long terme ou à l'occasion d'un événement particulier ?

- Si OUI : où en est le projet ? Avez-vous déjà défini une politique d'archivage ou arrêté des critères de sélection ?

- Si NON : ce projet vous semble-t-il d'un intérêt limité ou bien les obstacles vous ont-ils paru trop importants ? des obstacles de quel ordre ?

2/ Votre établissement a-t-il envisagé une collecte de sites régionaux ?

- Si OUI : où en est le projet ? Avez-vous déjà défini une politique d'archivage ou arrêté des critères de sélection ?

- Si NON : ce projet vous semble-t-il d'un intérêt limité ou bien les obstacles vous ont-ils paru trop importants ? des obstacles de quel ordre ?

3/ Connaissez-vous des BM françaises (ou étrangères) qui ont mis en oeuvre ou s'apprêtent à lancer un projet de ce type ?

4/ Selon vous, quel intérêt pourrait représenter un archivage du web régional pour vos lecteurs ?

5/ Quels types de sites web vous paraissent de nature à devoir faire l'objet d'une capture pour conservation à la BM ?

6/ Dans quels domaines principaux pourrait s'exercer la capacité d'expertise des bibliothécaires dans un projet de sélection de sites régionaux ?

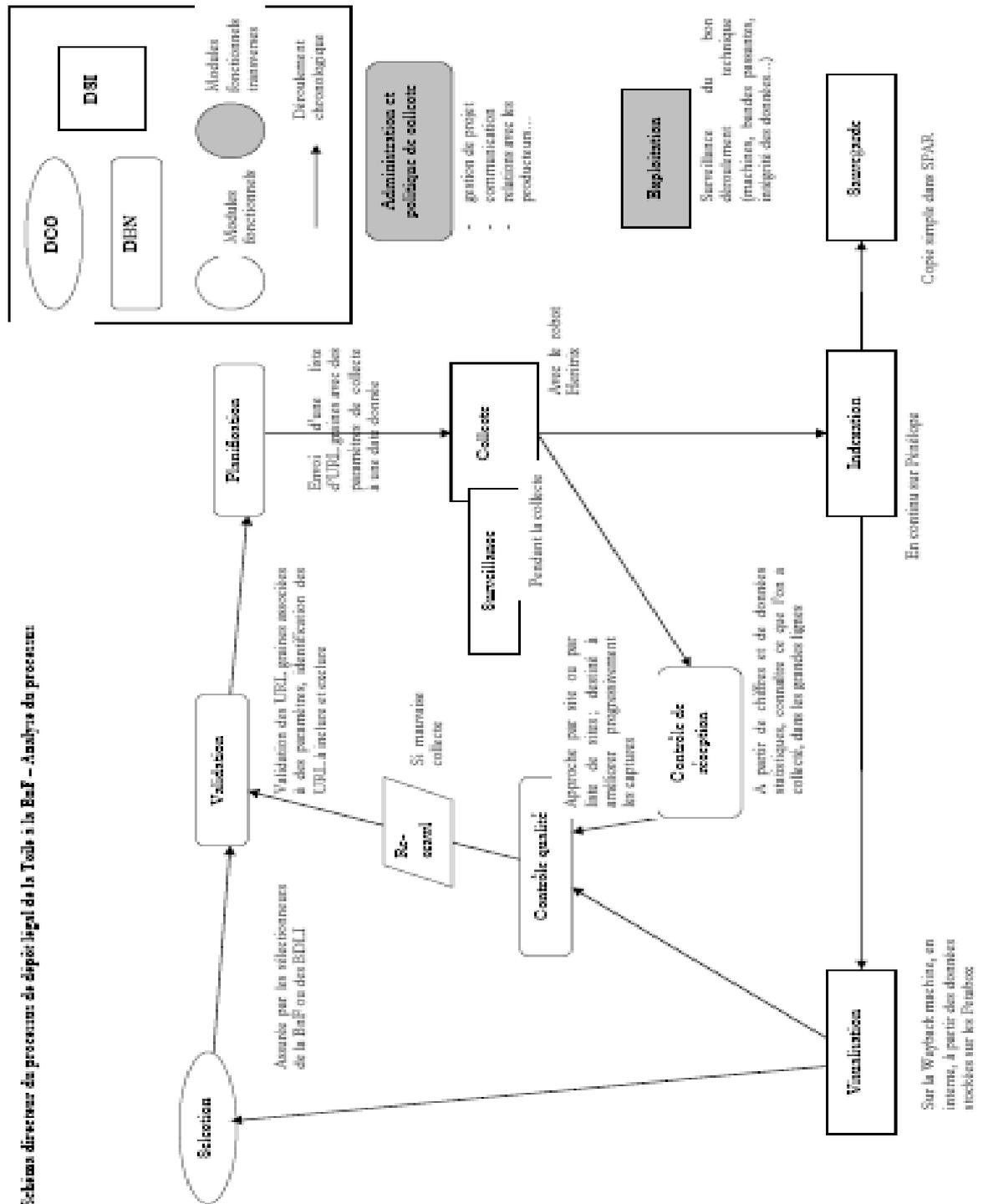
7/ Quel mode de restitution vous semblerait le plus adapté pour votre public ?



# Annexe 2

## SCHEMA DIRECTEUR DU PROCESSUS DE DEPOT LEGAL DU WEB A LA BNF.

(extrait de « Elections 2007 : Bilan de l'organisation du projet », Rapport BnF BnF-ADM-2007-031419-01, p. 22 ; reproduit avec l'aimable autorisation de Gildas Illien)





## Annexe 3

### LISTE D'ADRESSES U.R.L. DU WEB FOREZIEN DONNEES A TITRE INDICATIF

<b>INFORMATIONS GENERALES / ACTUALITES</b>	
<a href="http://www.forez-info.com">www.forez-info.com</a>	Site Forez-Info
<a href="http://www.zoom42.fr/alaune/UNE_index.asp">http://www.zoom42.fr/alaune/UNE_index.asp</a>	Zoom 42 (infos, annonces, etc)
<b>HISTOIRE LOCALE / PATRIMOINE</b>	
<a href="http://forezhistoire.free.fr">http://forezhistoire.free.fr</a>	Site Forez Histoire
<a href="http://www.ladiana.com/">http://www.ladiana.com/</a>	Site de La Diana (Montbrison)
<a href="http://cths.fr/an/liste.php?sc=ss&amp;nom=&amp;region=24&amp;domaine=0&amp;periode=0&amp;page=1&amp;x=133&amp;y=7">http://cths.fr/an/liste.php?sc=ss&amp;nom=&amp;region=24&amp;domaine=0&amp;periode=0&amp;page=1&amp;x=133&amp;y=7</a>	Fiches de présentation des sociétés savantes (Site du Comité des Travaux Historiques et Scientifiques)
<a href="http://saintchamond.canalblog.com/">http://saintchamond.canalblog.com/</a>	Site dédié à Saint-Chamond
<a href="http://lecorbusier.ville-firminy.fr/">http://lecorbusier.ville-firminy.fr/</a>	Site de la ville de Firminy – Patrimoine Le Corbusier
<a href="http://www.cheminferhautforez.com/">http://www.cheminferhautforez.com/</a>	Association de préservation d'une voie ferrée
<b>ADMINISTRATIONS</b>	
<a href="http://www.ia42.ac-lyon.fr/">http://www.ia42.ac-lyon.fr/</a>	Inspection académique - Loire
<a href="http://www.archives.saint-etienne.fr/site/index.asp">http://www.archives.saint-etienne.fr/site/index.asp</a>	Archives de Saint-Etienne
<a href="http://www.loire.fr/display.jsp?id=c_587862">http://www.loire.fr/display.jsp?id=c_587862</a>	Archives de la Loire (Site du Conseil général)
<a href="http://www.loire.fr/display.jsp?id=ci_527977">http://www.loire.fr/display.jsp?id=ci_527977</a>	Lettre d'informations du Conseil général de la Loire
<a href="http://www.agglo-st-etienne.fr">http://www.agglo-st-etienne.fr</a>	Agglomération de Saint-Etienne
<a href="http://www.tl7.fr">http://www.tl7.fr</a>	TL 7 – Télévision locale de Saint-Etienne
<a href="http://www.asse.fr">http://www.asse.fr</a>	Association sportive Saint-Etienne
<a href="http://www.saint-etienne.aeroport.fr">http://www.saint-etienne.aeroport.fr</a>	Aéroport de Saint-Etienne
<a href="http://www.saint-etienne.cci.fr">http://www.saint-etienne.cci.fr</a>	Chambre de commerce et d'industrie de la Loire
<a href="http://www.stas.tm.fr">http://www.stas.tm.fr</a>	Transports urbains de Saint-Etienne
<a href="http://www.avocat-saint-etienne.com">http://www.avocat-saint-etienne.com</a>	Ordre des avocats de Saint-Etienne
<a href="http://www.golf-st-etienne.com">http://www.golf-st-etienne.com</a>	Golf de Saint-Etienne
<a href="http://www.saint-etienne.urssaf.fr">http://www.saint-etienne.urssaf.fr</a>	URSSAF, Saint-Etienne
<b>MUSEES</b>	
<a href="http://www.citedudesign.com/main.php?lang=FR#1">http://www.citedudesign.com/main.php?lang=FR#1</a>	Cité du design
<a href="http://www.loire-musees.fr/">http://www.loire-musees.fr/</a>	Liste des musées de la Loire
<a href="http://musees-scolaires.saint-etienne.fr">http://musees-scolaires.saint-etienne.fr</a>	<i>Pages sur :</i>
	le Musée d'Art et d'Industrie

	le Musée de la Mine
	le Musée d'Art moderne
<a href="http://www.mam-st-etienne.fr/index.php">http://www.mam-st-etienne.fr/index.php</a>	Musée d'Art moderne, Saint-Etienne
<a href="http://www.vieux-saint-etienne.com">http://www.vieux-saint-etienne.com</a>	Société et Musée des Amis du Vieux Saint-Etienne
<a href="http://www.museedutissage.com">http://www.museedutissage.com</a>	Musée du tissage de Bussières
<a href="http://www.astronef.fr">http://www.astronef.fr</a>	Planétarium de Saint-Etienne
<a href="http://labastie.chez-alice.fr/">http://labastie.chez-alice.fr/</a>	Site personnel sur La Bastie d'Urfé
<a href="http://www.multitex.fr/bruneaux">http://www.multitex.fr/bruneaux</a>	Ecomusée des Bruneaux, Firminy
<a href="http://www.ville-st-jean-bonnefonds.fr/-La-Maison-du-Passementier-.html">http://www.ville-st-jean-bonnefonds.fr/-La-Maison-du-Passementier-.html</a>	Maison du Passementier, St-Jean-de-Bonnefonds
<a href="http://www.ecomusee-ussou-en-forez.fr">http://www.ecomusee-ussou-en-forez.fr</a>	Ecomusée d'Usson-en-Forez
<a href="http://www.mairie-roanne.fr/page.php?rub=04&amp;srub=01&amp;PHPSESSID=74e44012ed0384d719dd8250b4ea4365">http://www.mairie-roanne.fr/page.php?rub=04&amp;srub=01&amp;PHPSESSID=74e44012ed0384d719dd8250b4ea4365</a>	Musée de Roanne
<a href="http://www.ville-montbrison.fr/spip.php?rubrique88">http://www.ville-montbrison.fr/spip.php?rubrique88</a>	Musée d'Allard, Montbrison
<a href="http://www.amisdesartscharlieu.com">http://www.amisdesartscharlieu.com</a>	Société des amis des arts de Charlieu ; Musée de la Soierie et Musée hospitalier
<a href="http://www.pilat-tourisme.fr/Maison-des-tresses-et-lacets-a-La.html">http://www.pilat-tourisme.fr/Maison-des-tresses-et-lacets-a-La.html</a>	Parc naturel du Pilat ; Maison des tresses et lacets
<b>CENTRES CULTURELS / CENTRES DE SPECTACLES</b>	
<a href="http://www.comedie-de-saint-etienne.fr">http://www.comedie-de-saint-etienne.fr</a>	Comédie de Saint-Etienne
<a href="http://www.zenith-saint-etienne.fr">http://www.zenith-saint-etienne.fr</a>	Zenith de Saint-Etienne
<a href="http://www.andrezieux-boutheon.com/act">http://www.andrezieux-boutheon.com/act</a>	Ville d'Andrezieux-Boutheon ; Théâtre du Parc
<a href="http://www.theatredespenitents.fr">http://www.theatredespenitents.fr</a>	Théâtre des Pénitents, Montbrison
<a href="http://www.defrichecompagnie.org">http://www.defrichecompagnie.org</a>	Compagnie Défriche et Théâtre de l'Imprimerie ; Rive-de-Gier
<a href="http://theatreverso.free.fr">http://theatreverso.free.fr</a>	Théâtre Le Verso, Saint-Etienne
<a href="http://www.saint-etienne.fr/index.php?module=cms&amp;action=getContent&amp;id_heading_object=1040">http://www.saint-etienne.fr/index.php?module=cms&amp;action=getContent&amp;id_heading_object=1040</a>	Opéra-Théâtre de Saint-Etienne
<a href="http://biennalesaint-etienne.citedudesign.com">http://biennalesaint-etienne.citedudesign.com</a>	Biennale du design, Saint-Etienne
<a href="http://www.centre-culturel-laricamarie.com/medias/000_Push/img/Plaque_CCLR_0809.pdf">http://www.centre-culturel-laricamarie.com/medias/000_Push/img/Plaque_CCLR_0809.pdf</a>	Centre culturel de La Ricamarie
<b>NATURE, ENVIRONNEMENT</b>	
<a href="http://www.frapna.org/loire/content/view/full/149/85/">http://www.frapna.org/loire/content/view/full/149/85/</a>	Fédération Rhône-Alpes de Protection de la nature sur l'Ecopôle du Forez
<a href="http://www.parc-livradois-forez.org/">http://www.parc-livradois-forez.org/</a>	Parc naturel du Livradois-Forez
<a href="http://wiki.cerdd.org/tiki-index.php?page=Parc%20Naturel%20RC3%A9gional%20du%20Livradois%20Forez">http://wiki.cerdd.org/tiki-index.php?page=Parc%20Naturel%20RC3%A9gional%20du%20Livradois%20Forez</a>	Centre ressource du développement durable (groupement d'intérêt public). Web collaboratif environnemental.
<a href="http://blog.parc-livradois-forez.org/index.php?Developpement-">http://blog.parc-livradois-forez.org/index.php?Developpement-</a>	Blog du Parc Livradois-Forez consacré au nouveau projet pour le parc (atelier

<a href="#">durable</a>	citoyen)
<a href="http://etablissement-ecoresponsable.rhonealpes.fr/IMG/pdf/cpte_rdu_19_03_08.pdf">http://etablissement-ecoresponsable.rhonealpes.fr/IMG/pdf/cpte_rdu_19_03_08.pdf</a>	Démarche développement durable du Lycée du Forez (Feurs)
<a href="http://www.parc-naturel-pilat.fr">http://www.parc-naturel-pilat.fr</a>	Parc naturel régional du Pilat
<a href="http://blog.parc-livradois-forez.org/index.php?2008/06/06/69-atelier-citoyen-de-sylviculteurs-du-livradois-forez">http://blog.parc-livradois-forez.org/index.php?2008/06/06/69-atelier-citoyen-de-sylviculteurs-du-livradois-forez</a>	Parc Livradois-Forez, atelier citoyen de sylviculteurs
<b>ENSEIGNEMENT SUPERIEUR</b>	
<a href="http://www.institutdesign.com">http://www.institutdesign.com</a>	Ecole supérieure d'art et de design de Saint-Etienne
<a href="http://portail.univ-st-etienne.fr/">http://portail.univ-st-etienne.fr/</a>	Université Jean-Monnet, Saint-Etienne
<a href="http://www.iut.univ-st-etienne.fr/site/lance.htm">http://www.iut.univ-st-etienne.fr/site/lance.htm</a>	Institut Universitaire Technologique de Saint-Etienne
<a href="http://www.emse.fr/index.php">http://www.emse.fr/index.php</a>	Ecole des Mines, Saint-Etienne
<a href="http://www.st-etienne.archi.fr">http://www.st-etienne.archi.fr</a>	Ecole nationale supérieure d'architecture de Saint-Etienne
<a href="http://www.enise.fr">http://www.enise.fr</a>	Ecole nationale d'ingénieurs de Saint-Etienne
<a href="http://www.esc-saint-etienne.fr">http://www.esc-saint-etienne.fr</a>	Ecole Supérieure de Commerce de Saint-Etienne
<b>AIRE LINGUISTIQUE</b>	
<a href="http://patois.vivant.free.fr/images/patoisvivantlexique.pdf">http://patois.vivant.free.fr/images/patoisvivantlexique.pdf</a>	Patois des Monts du Forez : lexique collecté depuis les années 1970
<a href="http://patois.vivant.free.fr/patois.htm">http://patois.vivant.free.fr/patois.htm</a>	Site sur la Patois des Monts du Forez
<a href="http://www.w.rene-merle.com/article.php3?id_article=23">http://www.w.rene-merle.com/article.php3?id_article=23</a>	Site personnel : recherches historiques et sociolinguistiques sur l'occitan et le francoprovençal
<a href="http://www.lexilogos.com/forezien_dictionnaire.htm">http://www.lexilogos.com/forezien_dictionnaire.htm</a>	Dictionnaire forézien de Louis Gras (1863), grammaire, textes bilingues, parler de Saint-Etienne
<a href="http://corpusdelaparole.culture.fr/spip.php?article62">http://corpusdelaparole.culture.fr/spip.php?article62</a>	Enregistrements avec transcription
<a href="http://vieux.saint.etienne.club.fr/francoprovençal-arpitan/tintinparlegaga1.pdf">http://vieux.saint.etienne.club.fr/francoprovençal-arpitan/tintinparlegaga1.pdf</a>	Fichier sur le francoprovençal
<a href="http://www.arpitania.eu">http://www.arpitania.eu</a>	Site sur l'arpitan
<a href="http://www.univ-catholyon.fr/1205930136892/0/fiche_article/&amp;RH=1188919604365">http://www.univ-catholyon.fr/1205930136892/0/fiche_article/&amp;RH=1188919604365</a>	Bibliographie de l'Institut P. Gardette, Université catholique de Lyon