

Conserver les sites Internet en art contemporain : **enjeux et difficultés des collectes du web** **à la bibliothèque nationale de France**

Introduction : historique de la collecte et cadre légal

Dès 2001 la BnF a entamé, dans le cadre de sa mission de dépôt légal national, des expériences de collecte du web français.

Ces premières expériences étaient encore artisanales si l'on peut qualifier ainsi des manipulations faites avec Internet. Il s'agissait de contacter les auteurs de sites et de leur en demander une copie, soit par cassette à cette époque, par disque numérique ou par mail, à leur convenance.

Ces premiers essais ont déterminé la création d'un service spécifiquement consacré au dépôt légal du web, rattaché au département du Dépôt légal qui concerne tous les types de supports. Ce service réunit aujourd'hui 5 bibliothécaires, auxquels il faut rajouter quatre ingénieurs du département des Systèmes d'information.

La collecte s'effectue de deux façons :

- Une collecte dite « collecte large » sur l'ensemble des sites en .fr
- Une collecte intellectuelle faite par des spécialistes dans chacune des disciplines du fonds encyclopédique de la BnF dite « collecte ciblée ».

Les deux types de collecte sont réalisées grâce au robot Heritrix.

Ces collectes ciblées ont pour but de repérer des sites intéressants à préserver à tout prix que la collecte large n'aurait pas pris en compte, ce repérage intègre des données d'urgence (événement ponctuel comme un festival, évolution ou fermeture d'un site).

Depuis le 1^{er} août 2006, la loi DADVSI (droit d'auteur et droits voisins dans la société de l'information) autorise la collecte automatique en ligne des sites français en faisant une exception au droit d'auteur et désigne pour cette mission de conservation la BnF (ainsi que l'INA-Institut national de l'audiovisuel pour la radio et la télévision). Son décret d'application, publié en décembre 2011 ouvre de

nouvelles perspectives à la collecte, avec la possibilité de réclamer les codes d'accès aux auteurs des sites sur inscription et les données techniques qui empêchent le passage des robots collecteurs. La collection constituée représente aujourd'hui 250 teraoctet. Elle a été complétée en 2006 par l'achat à Internet Archive d'archives rétrospectives de 1996 à 2003 pour les sites en .fr

1. Une politique documentaire pour le web :

Le département littérature et art qui a été l'un des tout premiers à travailler sur ce projet, a établi dès 2006, lorsque le dispositif légal a été mis en place, une politique documentaire.

Largement inspirée par la nouveauté du support Internet, guidée par l'existant, elle a cherché à définir les spécificités et les opportunités documentaires offertes par ce nouveau media, tout en essayant de réfléchir aux limites à poser face à un ensemble aussi vaste.

Elle a également été très critique et précise dans ses remarques, faites au fil des expérimentations, dans la mesure où les sites d'art, particulièrement d'art contemporain, se sont révélés souvent difficiles à collecter.

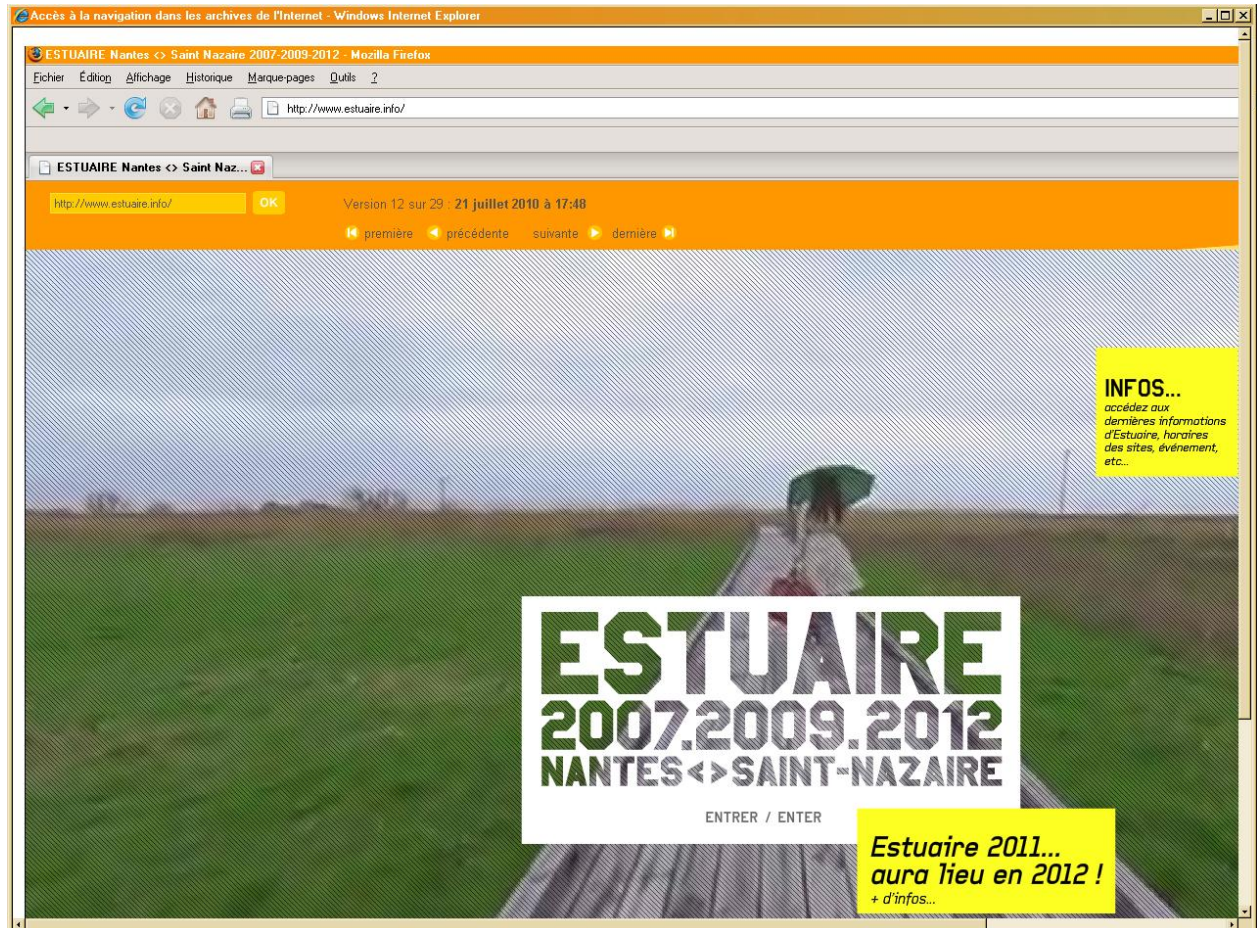
Ainsi il s'agissait de

- Recueillir une richesse documentaire exceptionnelle pour l'art :
 - Grâce à Internet on obtient une facilité de consultation inégalée. Il est possible d'inventorier des objets visuels très nombreux et de mettre facilement à jour ces inventaires, ceci pour un coût dérisoire.
 - Des œuvres non répertoriées ni éditées sur des supports papiers sont enfin visibles. Les sites des galeries fournissent des informations biographiques sur leurs artistes.
 - La toile est un espace de liberté pour les artistes émergents ou dissidents et pour des formes d'art en marge, un outil de diffusion souple et maîtrisable.
- Archiver l'actualité de l'art et ses manifestations éphémères :
Expositions, festivals, œuvres en ligne, œuvres éphémères, sites d'expositions temporaires, programmes des musées, événements artistiques, biennales, colloques. Et aussi les sites d'information, revues, réseaux, blogs, tous supports permettant de rendre perceptible et de documenter le monde contemporain.

- Couvrir toutes les disciplines qui sont celles de l'art et de l'histoire de l'art : du graffiti à la restauration d'art.

- Partager la récolte de la création contemporaine avec d'autres départements de la BnF qui traditionnellement acquièrent des œuvres par type de support : le département de l'Audiovisuel pour les œuvres de création multimédia

et en ligne, le département des Estampes et de la Photographie, et celui d'Histoire du livre pour les livres d'artistes et le graphisme.



<http://www.estuaire.info> (Archives de l'Internet, 21 juillet 2010)

Trois grands axes documentaires :

La création contemporaine :

- Les artistes

De nombreux artistes ont leurs sites personnels, pour certains cependant, ceux qui n'utilisent pas le media numérique, c'est assez récent, par exemple depuis 2004 pour Daniel Buren, bien qu'il ait réservé un nom de site depuis 2001 ; les artistes se regroupent souvent en réseaux, pour des motifs esthétiques, politiques, techniques ou simplement logistiques. Des artistes du Net ou qui créent des oeuvres interactives se regroupent dans des sites de laboratoires, des collectifs d'artistes ou des revues en ligne.

Les artistes marginaux comme les graffiteurs trouvent également avec Internet un moyen de diffusion hors des circuits officiels.

- Le monde artistique

Des sites essentiels sont ceux des galeries, sources précieuses d'informations biographiques sur les artistes. Mais sont aussi collectés les réseaux d'historiens, de chercheurs et de curateurs, les listes de discussion, les sites concernant la formation des artistes comme les programmes des écoles d'art et des centres de formation.

L'actualité

Nous préservons les sites d'expositions temporaires, les expositions en ligne, rares mais toujours intéressantes car ce sont des supports uniques, les visites virtuelles de musées, d'événements, les biennales, les colloques.

Les sites d'information, les revues, la télévision, les réseaux d'information, les blogs, tous supports permettant de rendre perceptible l'activité créatrice et de documenter le monde contemporain.

Les ressources documentaires

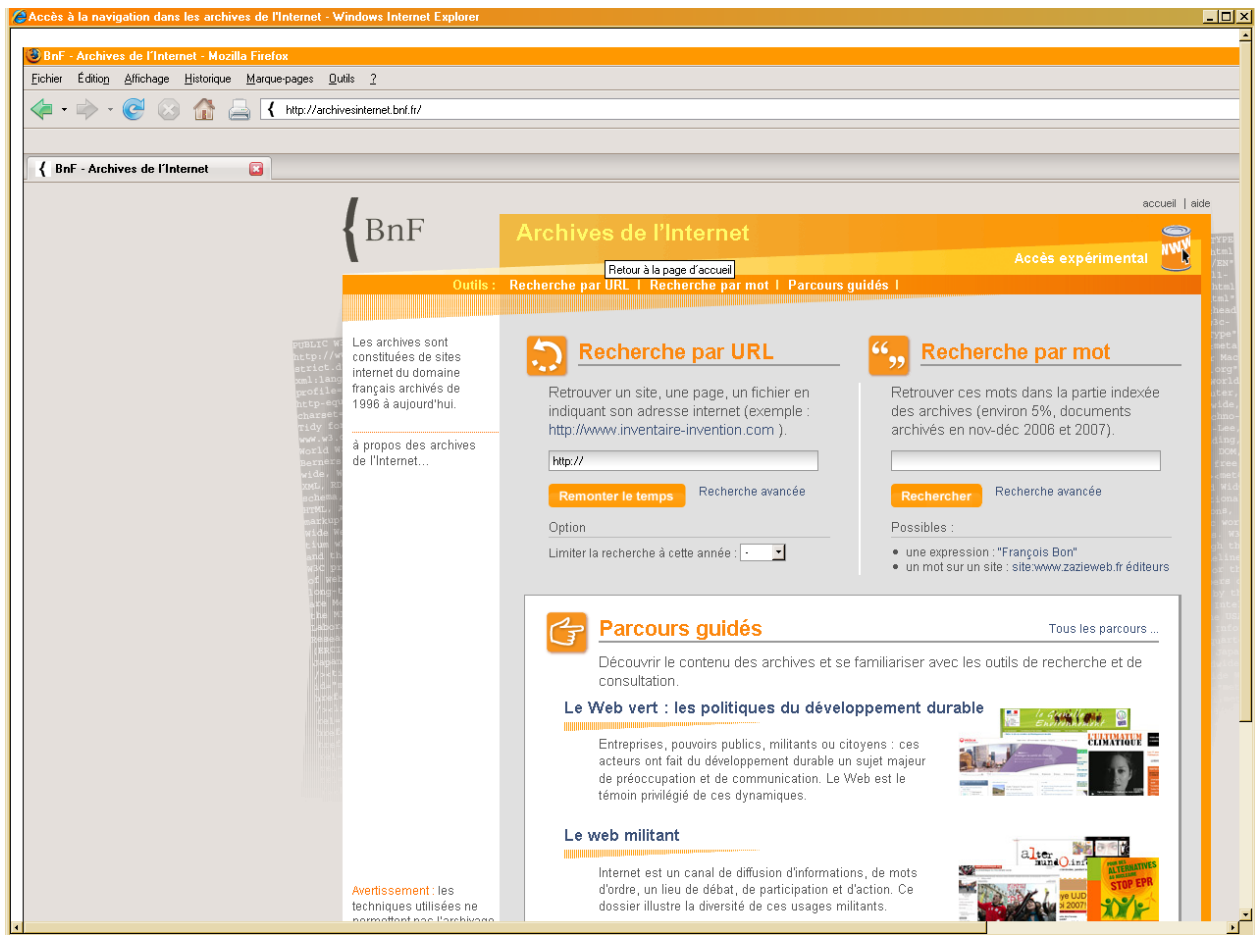
Il s'agit de la production éditoriale, les sites concernant le patrimoine, les inventaires d'œuvres d'art, le marché de l'art et les maisons de vente d'art, les sites des commissaires priseurs.

Les bases de données bibliographiques ne sont pas récoltées de même que les bases d'images, dont l'accès par un formulaire de recherche n'est pas adapté au robot collecteur que nous utilisons.

2. L'interface de consultation :

Consultables uniquement dans la bibliothèque de recherche accessible sur accréditation afin de préserver les droits d'auteur en cours, les « Archives de l'Internet » proposent deux moyens d'accès : une recherche simple par URL et des sélections thématiques éditorialisées.

L'art y est présent à travers un dossier sur les blogs (S') écrire en ligne : journaux personnels et littéraires. Il s'agit d'une sélection de blogs d'art dont certains ne sont plus en ligne ou ne donnent plus l'accès qu'à leurs archives les plus récentes



(Archives de l'Internet : page d'accueil)

3. Les limites de la collecte du web spécifiques à l'art :

Les collectes ciblées dont il est question ici sont faites avec un robot qui copie en profondeur chaque site. On lui fournit l'adresse précise, avec des paramètres de fréquence (mensuel, annuel, 1 ou 2 fois par semaine) il collecte également les liens sélectionnés par les auteurs du site, ce qui permet de capturer son environnement culturel.

La qualité de ces collectes est ensuite contrôlée par les bibliothécaires qui ont établi ces listes.

Les collectes se sont révélées excellentes quand les sites étaient constitués de textes, d'images fixes et que leur architecture était semblable à celle d'une page statique. C'est notamment le cas des blogs.

Cependant plusieurs autres difficultés apparaissent dès que les sites comportent :

-Des menus déroulants

-Des animations évolutives qui proposent des choix de navigation, un robot ne pouvant pas faire de choix, son mode d'action est simple : il suit les liens qu'il rencontre à chaque page, mais certainement pas des métamorphoses plastiques chargées de sens.

-Des bandes déroulantes, des panoramas. Ces animations sont en général faites avec le logiciel Flash que le robot n'a pas intégré, le problème est connu et identifié dans plusieurs autres bibliothèques.

-Une animation trop longue en ouverture d'un site peut empêcher le robot de poursuivre la collecte du site en profondeur, et ne pas même être récupérée.

-La présence de programmes spécifiques qui empêchent la visite des robots et la capture de leur contenu sont installés par les auteurs des sites.

-Les vidéos sont récoltées de manière inégale, souvent du fait de la reconnaissance du lien vidéo par le robot.

Dans tous ces cas, on peut obtenir un vide à l'emplacement d'une image manquante, même si elle est fixe, et en présence d'animations, le site ne s'ouvre pas du tout, laissant l'internaute face à une page unie, blanche.

Paradoxalement des images peuvent être récoltées par d'autres chemins, car référencées par d'autres sites, dans ce cas le lien à partir du site d'origine peut ne pas être navigable, alors que l'image est bien enregistrée dans les archives.

Un autre problème est l'évolution de l'utilisation de Flash, des sites ont pu être bien collectés depuis le début car créés avec des logiciels simples, mais dès qu'ils ont modernisé leur apparence avec Flash, nous ne le capturons plus.

Conclusion

Si dans le domaine de l'art, les bilans des collectes ont pu être particulièrement médiocres à leurs débuts, l'évolution de la qualité est très nette, grâce à l'expérience acquise, à la patiente identification des difficultés techniques et à l'affinement au cas par cas des stratégies de collecte.

Si les éléments d'information textuels récoltés ne sont pas négligeables (légendes, explications sur le travail des artistes, les œuvres, les mouvements, les galeries, les institutions, etc.) il est bien évident que nous ne pouvons pas nous en contenter.

Des améliorations restent donc à faire dans de nombreux cas, mais les sites d'art de par leur richesse formelle constituent sur ce plan un corpus d'expérimentation qui se situe à la pointe de ces recherches. Si on peut pointer qu'une partie du web a déjà disparu, les auteurs des sites, conscients de cette volatilité ont rapidement archivé leurs états successifs, créant et protégeant ainsi leurs archives.

Pour autant depuis 10 ans les collectes se sont grandement améliorées ainsi que les performances des robots qui ne cessent de s'affiner, aussi nous pouvons espérer offrir au public dans un avenir proche des images qui seront le fidèle reflet de l'activité artistique de nos contemporains.

Sites présentés :

<http://www.documentsdartistes.org>

<http://www.space-invaders.com>

<http://www.danielburen.com>

<http://www.magasin-cnac.org>

<http://taniuchi.fr>

<http://www.missticinparis.com>

<http://www.domaine-chaumont.fr>

<http://www.centrepompidou-metz.fr>

<http://lunettesrouges.blog.lemonde.fr>

<http://annemalherbe.blogspot.fr>

Françoise jacquet, Bibliothécaire, secteur art, département littérature et art de la bibliothèque nationale de France, chargée de la collecte des sites Internet français en art