

# Le défi de l'archivage de la totalité du Net

## L'Internet suscite-t-il de nouvelles règles d'archivages?

Pour les Archives nationales de France, l'Internet n'impose pas d'arbitrages nouveaux: «*Les mêmes règles que pour le papier s'appliquent*», souligne Joël Poivre, conservateur. L'institution assure «*uniquement*» la conservation des archives publiques (celles qui ne font pas nécessairement l'objet d'une publication) selon deux critères: l'intérêt administratif et juridique d'une part, l'intérêt historique ensuite. Le contenu des sites publics – ministères, organismes publics, etc. – est donc de son ressort. «*Mais ce qui est mis aujourd'hui en ligne par les ministères existe déjà sous une autre forme, papier ou fichiers informatiques*», explique Joël Poivre. Autrement dit, les Archives en ont déjà la mémoire. Depuis 1986 en effet, les Archives de France ont constitué un stock de 5000 fichiers informatiques, placées au Centre des archives contemporaines (CAC), à Fontainebleau. On y trouve des données météorologiques, des données de l'Insee, les recensements de la population, les grandes enquêtes nationales... ou encore l'histoire des personnels du ministère de l'Agriculture. Autant d'informations tenues à disposition des historiens.

## Peut-on archiver l'intégralité du Web?

A la Bibliothèque nationale de France (BNF), on caresse l'idée d'archiver la totalité du Web, privé comme public. Les Suédois poursuivent d'ores et déjà ce but. Comment? Une directive du gouvernement prévoit que des robots ratissent l'ensemble du Net suédois et mémorisent, non pas son intégralité, mais au moins tout ce que les moteurs de recherches (du type Yahoo! ou Google) indexent. Les générations futures ou les chercheurs n'auront «plus qu'à» interroger des bases de données pour trouver l'information qui leur manque. Le tout-venant de l'Internet sera donc gardé en mémoire, classé par date, comme autant d'instantanés de l'état du Web à un moment donné. N'échapperont à cet archivage que des pages non répertoriées. En tapant dans cinquante ans, par exemple, «vacances + 1999» sur un moteur de recherche, un internaute suédois accéderait aux statistiques de l'époque du ministère du Tourisme, aux sites des tours opérateurs, mais aussi au petit film que Björn a tourné sur la plage de Vieux-bouquet l'année de ses vacances dans les Landes et qu'il a mis en ligne sur son site personnel. Même si tous ces contenus ont disparu du Net dans l'intervalle.

## Faut-il conserver les sites individuels et les forums?

«*Aujourd'hui, on aime regarder des vieilles cartes postales en noir et blanc*, argumente Catherine Lupovici, responsable de la bibliothèque numérique de la BNF. *On note le même goût pour les vieilles publicités: ce n'est pas toujours l'information la plus noble qui est recherchée.*» Mais elle pose peut-être la question du droit à l'oubli. Peut-on tenir à la disposition des générations futures le petit film de Björn gambadant en maillot de bain? «*Là encore, l'Internet ne pose pas de question nouvelles, mais il en modifie l'échelle*, souligne Catherine Lupovici. *Doit-on détruire les archives de Libération au prétexte qu'une personne interviewée dans une édition ancienne en regrette la teneur?*» On envisage donc d'archiver les propos des internautes sur les forums de discussion. Leurs prises de bec commentant la décision de justice française contre Yahoo! sur la vente en ligne d'objets nazis feront peut-être un jour l'objet d'une étude approfondie...

## Comment assurer la conservation?

Pour l'instant, on pratique l'«aspiration»: un robot balaie les sites et les liens vers lesquels ils renvoient. L'idée étant de garder à terme la possibilité de naviguer sur les sites, comme aujourd'hui. Mais les technologies évoluent vite. En l'an 2000, on peut regarder les graffitis de Pompéi mais comment faire pour que les données d'aujourd'hui soit consultables dans le futur? Disques, disquettes, CD-Rom feront sûrement un jour office de fossiles. Certains ont pensé à créer des musées où l'on conserverait les machines en état de fonctionner comme de leur temps... Mais le coût de l'entretien des bécanes, le problème des pièces de rechange et celui de la formation du personnel aux méthodes antiques marque les limites de cette proposition. Les archivistes ont donc choisi de conserver les informations dans un format informatique «*le plus à plat possible*». Et les font glisser d'un support à un autre, au gré des évolutions. C'est le principe de la «migration». D'autres lui préfèrent la méthode de l'«émulation»: une machine moderne pourrait simuler les comportements de celles du passé. Mais l'émulateur universel est encore un fantasme. Les 5000 fichiers conservés à ce jour par les Archives nationales de France le sont sur des cartouches. Il y a vingt ans, elles auraient tenu sur 3000 bandes magnétiques. Puis sur 40 disques optiques. Aujourd'hui sur seulement 10 cartouches. Mais que se passe-t-il si l'on perd une cartouche? L'histoire des supports d'information prouve que plus on a augmenté la densité d'informations contenues sur un support, plus on a fragilisé ce support. C'est toute la différence entre une pierre taillée et une disquette... ●

«*On aime regarder des vieilles cartes postales. Ce n'est pas toujours l'information la plus noble qui est recherchée.*»  
Catherine Lupovici, de la BNF.