

QUESTIONS OUVERTES

Faut-il archiver le Web ?

Ce qui est affiché sur la Toile change très vite. Un archivage périodique est donc nécessaire pour en garder la mémoire, avec des difficultés qui ne sont pas seulement techniques.

Gildas ILLIEN

Brewster Kahle, ingénieur américain diplômé du MIT, a inventé le premier système de publication sur l'Internet, WAIS [Wide Area Information Service]. En 1995, il rend visite à des amis travaillant à Palo Alto sur le grand moteur de recherche de l'époque, Alta Vista. Il y découvre que la quasi-totalité du Web d'alors, 16 millions de pages indexées, tient dans l'équivalent de deux distributeurs automatiques de café : le voilà qui enlace, littéralement, tout l'Internet !

Marqué par cette expérience et fasciné par le mythe de la Bibliothèque universelle d'Alexandrie, B. Kahle fonde l'année suivante Internet Archive, un organisme à but non lucratif qui se fixe pour objectif d'archiver tout le Web afin de conserver la mémoire numérique du monde et de la rendre accessible à tous. Son site, www.archive.org, basé à San Francisco, permet depuis 1997 de retrouver ce que les bibliothécaires appellent les incunables de l'Internet (par exemple les premières pages du site Web du quotidien *Le Monde*).

Les pouvoirs publics étant toujours un peu plus lents face à l'innovation, il faudra attendre août 2006 pour que le Parlement français vote la loi DADVISI qui institue un dépôt légal des « signes, signaux, écrits, images, sons ou messages de toute nature » publiés dans le cadre d'une communication au public par voie électronique – c'est-à-dire un dépôt légal de l'Internet. Celui-ci est confié à la Bibliothèque nationale de France (BnF), chargée de la collecte des sites français, à l'exception des sites des émetteurs de la radiotélévision qui relèvent, eux, de l'Institut national de l'audiovisuel (INA).

Cependant, les bibliothèques nationales – la BnF, mais aussi ses homologues de Grande-Bretagne, de Suède, du Danemark ou d'Australie – n'ont pas attendu la mise en place d'une législation pour engager des expérimentations visant elles aussi à mettre le Web en boîte à des fins patrimoniales. L'aventure de l'archivage du Web a débuté il y a plus de dix ans. Les ingénieurs pionniers ont d'abord vu le défi technique d'une telle entreprise. Aujourd'hui, alors que notre mémoire du monde numérique commence à faire défaut, beaucoup s'aperçoivent que cet archivage soulève aussi des questions de société : peut-on vraiment tout conserver ? À qui et à quoi un Web archivé pourra-t-il servir ?

Oubli ou mémoire ?

Google n'a que 15 ans, Wikipedia 10 ans, Facebook à peine 5 ans. Et nous ne savons pas ce qu'il restera en ligne, dans un an ou dix, des blogs et des réseaux sociaux ayant contribué à la propagation des révoltes en Tunisie et ailleurs. Faute de recul, il est difficile de savoir ce qui mérite d'être conservé et ce qui n'a pas été fait pour durer. Cette problématique n'est pas nouvelle, mais est centrale dans tout projet patrimonial. Les contours et les usages du patrimoine « numérique » (c'est-à-dire issu du Web, et non de la numérisation des livres) sont, de fait, encore difficiles à cerner.

Internet est un média éphémère, où l'instantanéité est la règle, mais nous y accumulons une quantité croissante de données numériques : photographies, conversations, enregistrements sonores et filmés, publications sur des blogs, des réseaux sociaux...

Avec le Web 2.0, ces contenus qui constituent des pans entiers de notre vie numérique ne sont pas seulement (voire pas du tout) conservés sur nos disques durs, mais directement sur la Toile.

Le déluge de données est tel que la difficulté n'est plus tant de trouver les informations que de les trier et de supprimer celles n'ayant pas d'utilité. À première vue, ce grand ménage semble se faire tout seul. Les webmasters évaluent la durée de vie moyenne des sites à deux ou trois ans : le Web s'évaporerait ainsi naturellement. Et au fond, l'idée que les données d'hier disparaissent à mesure que de nouvelles strates s'ajoutent serait plutôt rassurante : à quoi bon une mémoire numérique si les marées du Web peuvent, d'elles-mêmes, nous éviter de trier ?

Cathy Marshall, de la Société Microsoft Research, a étudié les comportements des Américains confrontés à leur propre mémoire numérique. Beaucoup estiment qu'il n'est pas utile de se préoccuper de leur conservation au motif que « d'autres » (« quelqu'un », Google, le Gouvernement fédéral, la Bibliothèque du Congrès...) s'en chargeraient déjà. Elle a même perçu chez certains un « amour du risque », pour qui la disparition volontaire ou accidentelle de leurs données aurait l'effet d'une libération – ce serait une page de sa vie que l'on pourrait enfin tourner : affranchi de son existence numérique, on peut tout recommencer à zéro.

Le principe même d'une mémoire numérique n'est-il pas fondamentalement dangereux ? Les grands moteurs indexeurs comme Google assureraient une mémoire totale, capable de conserver vos meilleurs souve-

nirs comme les pires et de ruiner votre réputation. Selon une récente enquête du CRÉDOC (Centre de recherche pour l'étude et l'observation des conditions de vie), les Français seraient préoccupés par la confidentialité et la protection des données sur Internet, et l'opinion se montrerait très favorable au « droit à l'oubli numérique » cher au sénateur président de la CNIL (Commission nationale de l'informatique et des libertés), Alex Türk, et à l'ancienne secrétaire d'État à l'économie numérique, Nathalie Kosciusko-Morizet. En effet, 91 pour cent des personnes interrogées estiment que « les sites Internet devraient permettre à chacun d'effacer simplement les informations personnelles qui ont été communiquées ».

Le spectre de « Big Brother »

Dans cette actualité où la mémoire peut sembler à la fois superflue et dangereuse, on peut donc se demander pourquoi une institution publique, telle la BnF, s'évertue à archiver des échantillons du grand fatras du Web ; serait-elle capable de trouver une juste mesure entre le spectre de *Big Brother* et la perspective tout aussi angoissante d'un monde privé d'histoire ?

Les internautes manquent peut-être de distance historique sur leurs propres pratiques pour se préoccuper de questions longtemps réservées aux chercheurs et aux conservateurs du patrimoine. C. Marshall observe cependant que la « non-politique » de conservation pratiquée par la majorité d'entre nous permet, assez paradoxalement, de retrouver par hasard des données oubliées. On compare souvent cette pratique de la « redécouverte » des fichiers oubliés dans les tréfonds des répertoires à l'expérience magique « d'une boîte à chaussures dans laquelle on retrouverait, pêle-mêle, de vieilles photographies, des coupures de presse, des lettres d'amour ». Une image qui n'est pas sans rappeler, pour les plus anciens,

1. LES MAGASINS NUMÉRIQUES de la Bibliothèque nationale de France, qui contiennent les archives du Web français stockées sur disques et sur bandes magnétiques.

