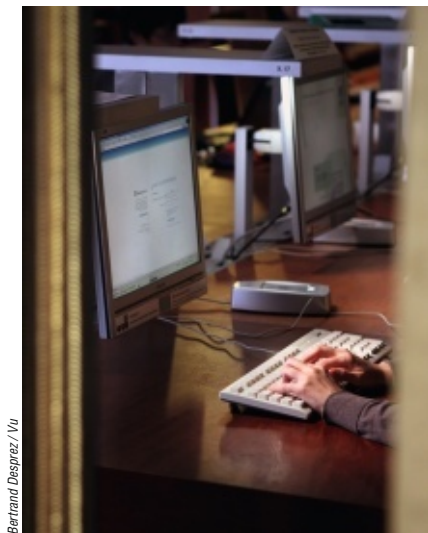




Bertrand Desprez / Vu

3. L'ACCÈS AUX ARCHIVES DU WEB est souvent strictement encadré en raison des données à caractère personnel qu'elles contiennent. À la BnF, cet accès est ainsi réservé aux chercheurs. En Scandinavie, les archives Web ne sont pas accessibles. En revanche, celles du site privé américain www.archive.org sont disponibles en ligne.

✓ SUR LE WEB

Le consortium IIPC :
<http://netpreserve.org/>

Dépôt légal des sites Web
à la Bibliothèque nationale
de France :

http://www.bnf.fr/fr/professionnels/depot_legal/a_dl_sites_web_mod.html

G. Illien, Le dépôt légal de l'Internet
en pratique : les moissonneurs
du Web, *Bulletin des Bibliothèques
de France*, n° 6, pp. 20-27, 2008,
disponible sur :

<http://bbf.enssib.fr/consulter/bbf-2008-06-0020-004>

Bibliographie établie par la BnF :
http://bibnum.bnf.fr/conservation/bibliographie_dl_web.pdf

CRÉDOC, La diffusion
des technologies de l'information
et de la communication
dans la société française,
décembre 2010
(rapport disponible sur
<http://www.credoc.fr/>).

discours, vidéos et autres documents numériques qui cristallisent les grands enjeux de société d'une époque. Le site de campagne du candidat Jacques Chirac archivé en 2002 (www.chiracaveclafrance.net, disparu depuis), où l'on trouve par exemple son appel aux Français le soir du premier tour de l'élection du 21 avril 2002, y fait figure de pièce de collection...

La BnF suit la vie des mouvements sociaux en ligne et conserve les sites des organisations syndicales, des grandes associations et des grandes causes, mais aussi des entreprises cotées au CAC 40 dont elle archive notamment tous les rapports annuels. Elle s'intéresse aux arts et à l'écriture littéraire. L'une de ses collections les plus fascinantes est faite de journaux personnels en ligne et de sites d'écrivains connus ou non. C'est une sorte de salon littéraire virtuel où l'on découvre les traces des premiers blogueurs (qui étaient des informaticiens) et de leurs suivants : policiers, infirmières, agriculteurs ou caissières qui, un jour, ont préféré le clavier à la plume pour raconter leur vie ou livrer leurs observations. Ces récits de vie numériques, qui côtoient les premiers brouillons et essais en ligne d'écrivains de renom, forment une galerie inédite ; pourquoi l'histoire de la littérature devrait-elle s'arrêter à Casanova, Proust ou Flaubert ?

Le robot est également programmé pour réagir aux urgences et aux grands événements imprévus. Les bibliothécaires suivent l'actualité et sont parfois confrontés à des choix difficiles. Ceux de la BnF ont, par le passé, pris des initiatives de collecte spontanée, jusque dans la rue, dont les historiens leur sont redevables aujourd'hui, comme ce fut le cas pendant l'Occupation ou Mai 68.

Ainsi, en décembre 2010, la BnF décida d'archiver l'intégralité du site controversé *Wikileaks* : l'un des serveurs miroirs étant hébergé chez OVH à Roubaix, le site était juridiquement assujéti au dépôt légal français, fondé depuis cinq siècles sur la relation au territoire national. Il ne s'agissait pas de prendre parti, mais de s'assurer que les « archinautes » du futur auraient les moyens de retrouver trace d'une affaire et d'un débat planétaires, qui seront probablement déterminants pour l'évolution de la régulation des

usages et de la liberté d'expression sur Internet. L'initiative était pertinente puisque, au même moment, le gouvernement américain interdisait l'accès au site *Wikileaks* depuis les agences fédérales américaines, y compris à la Bibliothèque du Congrès.

En prise avec l'actualité

Quelques semaines plus tard, en partenariat avec *Internet Archive*, les robots de la BnF partaient moissonner les sites de Tunisie et d'ailleurs qui ont permis de suivre les révoltes d'Afrique du Nord et du Moyen-Orient. On peut se réjouir, pour une fois, que la France ne soit pas la plus démunie juridiquement et techniquement pour se protéger, face à ce type d'événements, contre l'oubli.

Entre droit à l'oubli et devoir de mémoire numérique, on voit ainsi s'opposer des arguments convaincants de part et d'autre. Idéalement, un archivage qui ne conserverait du Web que sa partie publique, filtrant toute donnée personnelle, constituerait évidemment le compromis le plus acceptable entre la protection des individus et la conservation d'un patrimoine commun. Ce scénario est toutefois impossible à mettre en œuvre dans l'état actuel du droit et des technologies : même si les collectes de la BnF ne représentent qu'une part infime de la totalité des données publiées en ligne, pour rester représentatives elles doivent être effectuées à une échelle suffisamment grande, ce qui implique une large part d'automatisation.

Or identifier et filtrer les données personnelles à cette échelle n'est pas possible, pour des raisons économiques et informatiques, mais aussi parce qu'il est difficile d'apprécier le caractère personnel des informations. Les pratiques d'expression en ligne ont rendu très poreuses les lignes de partage habituelles entre espace public et privé, entre publication et communication interpersonnelle. De ce point de vue, les robots d'archivage ne sauraient faire la part des choses, même s'ils respectent les espaces privés lorsqu'ils sont clairement signalés comme tels (par exemple, le robot n'a pas d'amis sur *Facebook* et ne peut s'introduire dans un réseau social à moins qu'il ne soit déclaré public et visible par tous).