

Une collaboration entre l'APA et la BnF : L'archivage des journaux personnels en ligne

Bernard Massip, *La Faute à Rousseau* n° 47, février 2008

Il se constitue aujourd'hui sur internet sous forme de sites personnels ou de blogs un nouveau champ de l'expression autobiographique. Jour après jour les pages déposées construisent un vaste réservoir de récits de vies et de documents significatifs de l'état de la société, des idées, des modes relationnels et de l'air du temps, qui seront pour la postérité de précieux témoignages sur notre époque.

Mais cet océan de paroles personnelles est fragile et instable. Un document en ligne tel jour peut être devenue inaccessible six mois plus tard, soit que son auteur l'ait retiré, soit que les liens qui y mènent aient été brisés, soit que le serveur qui l'héberge ait disparu. Et cela dans l'histoire courte du web d'aujourd'hui ! Qu'en sera-t-il à une échelle de temps historique, que restera-t-il des sites de la fin du 20^e siècle à l'orée du 22^e?

L'archivage d'internet est donc une nécessité pour préserver la mémoire de notre époque. Le législateur en est conscient qui a élargi par une loi de 2006 le dépôt légal aux sites web. Contrairement à ce qu'il en est pour les autres supports, les éditeurs ou auteurs ne sont pas tenus de déposer eux-mêmes leur production. L'archivage s'effectue par des collectes automatisées réalisées par des robots que pilotent les institutions depositaires. Les blogs et autres journaux d'expression personnelle en ligne entrent dans le champ de ce dépôt légal et peuvent donc être collectés sans que les internautes concernés n'aient en rien à intervenir mais sans non plus qu'ils aient à donner leur accord.

La collecte exhaustive d'internet est impossible et il faut donc trouver les moyens d'un archivage le plus représentatif possible. Gildas Ilien, responsable du dépôt légal internet à la BnF a posé la problématique et les méthodes de cette collecte dans un récent article donné à notre revue (FAR n° 45, p 63). Le modèle retenu repose sur la complémentarité de diverses procédures, associant collectes larges à grande échelle et collectes ciblées et plus approfondies sur des sujets particuliers, réalisées éventuellement en collaboration avec des partenaires extérieurs.

La BnF a inclus la thématique des « écritures du moi » dans ses champs d'exploration et a souhaité sur ce thème pouvoir profiter de l'expertise de l'APA.

Cette collaboration s'est amorcée sans tarder dès le début de l'été. Concrètement nous avons transmis à la BnF une première liste d'une centaine de sites qui nous semblaient représentatifs de diverses tendances de l'expression personnelle en ligne. Nous nous sommes efforcés d'y inclure certains sites publics mais relativement discrets, peu fréquemment mis en lien, donc peu susceptibles d'avoir été inclus dans les collectes de masse.

Une première opération d'archivage a été effectuée courant août à partir de cette liste.

Nous avons rencontré l'équipe de la BnF courant novembre afin de faire un bilan de ce début de collaboration. Notre apport a été jugé très intéressant par nos interlocuteurs : Sur la centaine de sites répertoriés, une quarantaine seulement étaient déjà présents au titre de la collecte large et encore ne l'était-il la plupart du temps que de façon très lacunaire.

La collecte effectuée de façon ciblée est apparue dans l'ensemble satisfaisante, la plupart des sites récupérés l'ont été correctement et de façon assez complète. Lorsque des éléments manquent ce sont en général des éléments de mises en forme dont l'absence n'empêche pas la lecture des textes ni la navigation. De plus dans chacune de ces situations des hypothèses ont pu être faites sur la cause de ces lacunes permettant d'envisager des solutions. Certains sites à l'architecture particulièrement complexe (comme « Désordre » de Philippe De Jonckheere, voir FAR n°42, p 4 et n°45, p 30) posent des problèmes plus difficiles mais qui permettent aux ingénieurs de la BnF d'améliorer les procédures de collecte pour la rendre efficace dans toutes les situations.

Cette collecte aussi justifiée soit-elle pose cependant certains problèmes tenant à la nature de ce qu'est une publication sur internet et surtout de la conscience que leurs auteurs peuvent en avoir.

Il n'est pas certain en effet que l'internaute qui dépose un billet sur un site à destination de membres de sa famille ou de quelques amis proches ou bien au contraire à destination de lointains lecteurs anonymes ait tout à fait conscience de procéder à une « publication ». Il pourrait être étonné que ses modestes pages fassent l'objet d'un archivage. Certains ne manquent pas d'ailleurs d'évoquer alors l'ombre de Big Brother, venant épier les mots que tout un chacun produit derrière son écran et même les copier silencieusement pour les conserver dans une mémoire titanesque et qui pourrait être perçue comme potentiellement totalitaire.

Pendant la campagne du mois d'août un internaute attentif, Swami Petaramesh, a repéré le passage du robot de la BnF. Le robot, et c'est bien normal, n'opère pas clandestinement et s'identifie (pour qui toutefois le repère ce qui n'est pas à la portée de l'internaute lambda), indiquant qu'il « procède à une opération de collecte réalisée dans le cadre du dépôt légal d'internet ». L'internaute concerné s'en est étonné, ce qui a conduit à une discussion sur son blog (voir sur <http://petaramesh.org/> le billet du 27 août 2007) et par mail entre lui et les responsables du projet à la BnF. Sans remettre en cause l'intérêt du dépôt légal il s'interrogeait néanmoins sur sa pertinence et sa justification pour des sites à public volontairement restreint ou à vocation très temporaire.

Ainsi évoquait-il le blog de sa fille de 9 ans, « pouvant contenir accidentellement des infos personnelles ou des photos de famille que l'on ne souhaite pas voir archiver hors de tout contrôle. Ma fille tient son site pour une douzaine de copines de classe, pas pour la bibliothèque nationale ou je ne sais quel dépôt légal ».

De façon plus générale il serait intéressant de savoir comment l'écrivain en ligne perçoit son écriture du point de vue du rapport au temps et à la conservation. La considère-t-il comme essentiellement écriture de l'instant, écriture éphémère qui n'a de sens que par rapport aux dialogues immédiats qui s'instituent par les échanges de commentaires et qui n'a pas vocation à être archivée ? Ou bien la perçoit-il comme une trace de lui-même progressivement enrichie à mesure que s'accumulent les textes déposés, comme un « work in progress » qui finit par constituer une œuvre ? Il y a fort à parier que l'on trouverait parmi les blogueurs toutes une gradation d'attitudes sur ces questions. Certains revendiqueraient plus volontiers un droit à la mémoire alors que d'autres voudraient plutôt un droit à l'oubli. Mais qui sait aussi, si certains, après avoir voulu l'oubli, ne seraient pas heureux de retrouver des pages qu'ils croyaient disparues.

Il convient de rappeler qu'un site internet public, un blog public est par définition une publication. Son public peut être restreint en fait, il ne l'est pas en droit. La seule façon d'en limiter l'accès de façon effective est de le mettre sous mot de passe, réservé à des visiteurs identifiés. Il devient alors un écrit privé et n'est plus susceptible d'être versé au dépôt légal. Chacun doit en être conscient et réaliser à cette occasion qu'il peut être confronté à des visiteurs bien moins délicats que la BnF qui agit pour sa part dans le cadre d'une stricte déontologie et dans le but de réaliser sa mission de service public de conservation du patrimoine. Et l'APA est tout à fait dans son rôle lorsqu'elle contribue avec ses modestes moyens à préserver des témoignages qui sans cela auraient couru le risque d'être perdus.

Reste à établir les futures modalités de consultation de ces archives. Des garde-fous sont prévus afin d'éviter d'éventuelles diffusions larges que les auteurs n'ont ni souhaitées ni même envisagées. Il n'est pas question notamment que ces archives soient elles mêmes mises en ligne. Elles ne seraient consultables que sur place et sur accréditation.

L'APA par son action contribue ainsi à constituer, à l'ombre tutélaire de la BnF, une sorte de petite Grenette virtuelle. Mais une Grenette inconsciente d'elle-même. Car il s'agit là d'une action de simple transmission de notre part qui n'implique pas l'adhésion active des internautes concernés. Elle préserve des documents pour l'avenir et s'inscrit en ce sens tout à fait dans notre vocation patrimoniale. Mais on pourrait souhaiter aussi que certains de ces documents entrent aussi dans notre fonds d'Ambérieu qui accepte désormais d'accueillir et d'archiver des documents multimédias sur des supports tels que des cd-rom. Cela permettrait qu'ils soient lus dans nos groupes de lecture, qu'ils fassent l'objet d'un écho et d'un catalogage, bref que leur soit assuré ce petit supplément de visibilité que confère l'intégration dans le réseau de l'association. Nous ne manquerons pas de réfléchir dans les mois qui viennent aux méthodes qui seraient les plus adaptées pour développer pour nous-mêmes ces collectes d'un nouveau type.