

Le tournant du numérique pour le dépôt légal à la BnF

Le web a fait évoluer l'activité littéraire et autobiographique, et les livres et écrits numériques se multiplient aujourd'hui. Le système du dépôt légal fonctionnait relativement bien pour les publications imprimées, en dépit de quelques failles concernant les publications autoéditées ou éditées à compte d'auteur. Mais, au tournant du numérique, le dépôt légal devient plus complexe. De nouvelles missions se font jour pour la Bibliothèque nationale de France, qui a entamé un travail de repérage, de collecte, et d'adaptation de ses procédures pour accueillir dans ses collections ces œuvres littéraires d'un genre nouveau.

Les écrits numériques

Les formes que prennent aujourd'hui les écrits numériques sont extrêmement diverses, du livre numérique homothétique d'un livre imprimé aux multiples formes de la littérature en ligne. Internet fait bouger les frontières entre les écrits désignés comme littéraires ou pas et conduit à l'invention de nouvelles formes, plus courtes, fragmentaires, collaboratives, où s'invitent l'image et le son. Si certains blogs ou sites d'écrivains sont les avatars numériques des avant-textes que les bibliothèques conservaient jusqu'alors sous la forme de manuscrits ou d'archives personnelles, d'autres sont des œuvres à part entière mais qui, contrairement aux livres, se construisent et évoluent au fil du temps. Il convient de veiller dès maintenant à la préservation de ces œuvres pour les générations futures.

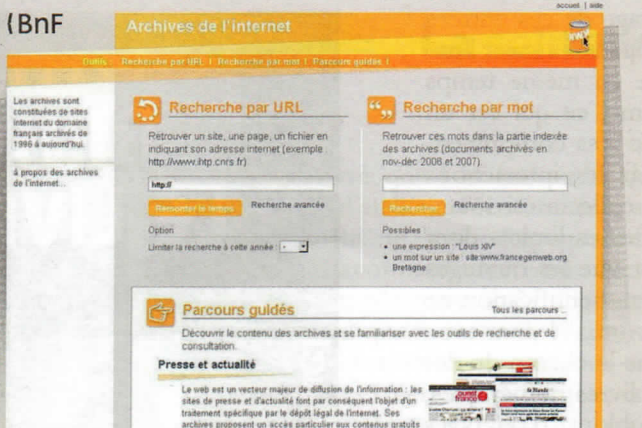
Les blogs et sites personnels de création « sont toujours susceptibles de disparaître ou de déménager ».
Images BnF



Le dépôt légal est l'obligation pour tout éditeur ou producteur de déposer chaque document édité, imprimé ou importé auprès de l'un des organismes dépositaires du dépôt légal, parmi lesquels la BnF. Il est obligatoire pour les livres dès lors qu'ils sont mis à la disposition d'un public qui excède le cercle de famille, à titre gratuit ou onéreux. Les livres autoédités ou édités à compte d'auteur sont aussi concernés. Les livres numériques le sont également, et si la diffusion d'un livre numérique coexiste avec une version sur support papier, un type de dépôt ne se substitue pas à un autre. Mais le dépôt s'avère plus difficile à mettre en œuvre pour les écrits et les livres numériques et passe pour le moment par les collectes automatiques du dépôt légal de l'internet.

L'idée de constituer une mémoire du web est née dans les années 1990 quand la toile a commencé à prendre de l'ampleur. En France les premières réflexions et expérimentations en 2000-2001 ont abouti à une loi en 2006. La BnF est depuis 10 ans chargée avec l'INA (à qui revient le dépôt légal des sites de radio et télévision) de constituer une mémoire du web français, baptisée dépôt légal pour s'inscrire dans une longue histoire, depuis l'instauration du dépôt légal des livres en 1537 par François 1^{er}, même s'il s'agit plutôt d'une archive.

Entreprendre de constituer une mémoire du web, même uniquement français, c'est un peu s'attaquer à l'océan à la petite cuiller ! Le dispositif choisi par la BnF évolue sans cesse pour tenter de suivre les évolutions rapides du web, et



Interface de recherche dans les archives du Web

repose sur une solution mixte qui mêle deux modes d'entrée : des collectes larges les plus automatisées possibles et des collectes ciblées plus fines mais limitées en taille. Les collectes larges sont réalisées une fois par an ; depuis 2007 elles sont lancées à partir de la liste de tous les sites en .fr fournie par l'AFNIC et des listes de quelques autres hébergeurs, soit environ 4,5 millions de sites lors de la dernière collecte à l'automne 2015. Ce mode de capture a toutefois des défauts : la profondeur de la capture est souvent insuffisante, les sites personnels et les blogs ne sont pas toujours collectés. C'est la raison pour laquelle a été ajouté un deuxième mode d'entrée qui a pour but de repérer mieux les sites qui pourraient échapper à la collecte large. Les collectes ciblées, plus limitées en nombre, permettent de constituer des archives plus complètes, plus profondes et plus fréquentes. Elles sont lancées à partir de listes de sites repérés par des bibliothécaires spécialistes de chaque discipline, ainsi que des universitaires ou des partenaires extérieurs, parmi lesquels Bernard Massip et Elizabeth Legros Chapuis pour l'APA. À ce jour près de 39 000 sites sont ainsi archivés via les collectes ciblées, parmi lesquels de très nombreux blogs et sites personnels. La BnF conserve ainsi des archives du web qui s'étendent de 1996 à aujourd'hui, soit plus de 668 téraoctets de données. Afin de protéger les auteurs, ces archives ne sont pas mises en ligne, leur consultation ne peut s'effectuer que dans les salles de Recherche de la BnF et de quelques bibliothèques partenaires, et aucune copie numérique n'est possible.

Les blogs et sites personnels de création littéraire entrent dans le champ de ce dépôt légal et peuvent donc être collectés sans intervention ni accord des internautes concernés. Leur collecte est d'autant plus importante que ces sites sont fragiles, éphémères, nomades, toujours susceptibles de disparaître ou de déménager. La veille documentaire concernant les blogs et les sites d'écrivains est assurée par l'équipe du Département Littérature et Art. La coopération avec l'APA pour la collecte des journaux personnels a également permis, depuis 2007, de collecter deux fois par an plus de 1000 sites et blogs. Grâce au dépôt légal du web, plusieurs milliers de sites et blogs d'écrivains ou d'écrivains français sont déjà entrés dans les Archives de l'internet et viennent y prolonger les collections imprimées. En outre chacun, écrivain ou pas, peut demander la collecte de son site à l'adresse depot.legal.web@bnf.fr, où arrivent plusieurs centaines de demandes de collecte par an.

Les livres numériques

En attendant la mise en production d'autres modes d'entrée, encore à l'étude, seules les collectes du web permettent à la BnF de remplir sa mission de dépôt légal des livres numériques. De fait, le dépôt légal du web permet d'ores et déjà de collecter de nombreux livres numériques qui sont en accès libre sur des sites généralistes ou sur les sites de leurs auteurs. Collectes larges ou ciblées et demandes de webmasters offrent un échantillon conséquent des livres numériques disponibles gratuitement sur le web français : une étude récente a permis de compter près de 45 000 epub dans les collections, et de constater que la BnF conserve dans les Archives de l'internet d'assez nombreux livres numériques devenus indisponibles sur le web aujourd'hui.

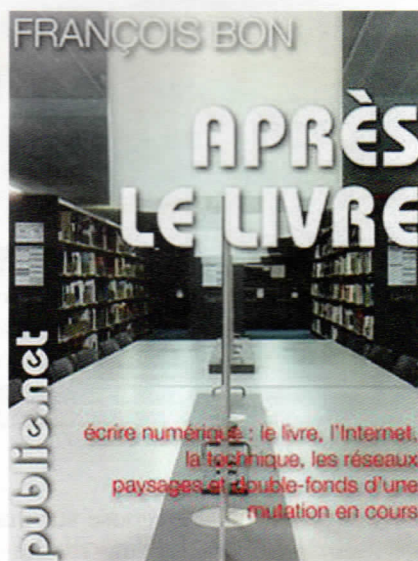
Vient d'ailleurs aussi d'être créée une collecte consacrée au Livre numérique, qui a pour objectif de sélectionner des sites représentatifs des différents acteurs de la chaîne : écrivains, technique du livre, édition, diffusion, bibliothèques, réception, analyse, sans omettre les nombreux blogs sur ces différents sujets.

Le principal avantage de ces entrées par collecte est leur capacité à saisir l'environnement du livre en même temps que le fichier lui-même, ce qui est un atout lorsqu'il est utile à sa compréhension, par exemple par les interactions entre livre et site. La collecte est toutefois confrontée à certaines limites dont la plus importante est que les ouvrages payants ou soumis à authentification ne sont pas collectés.

S'il est entendu que le champ du dépôt légal s'étend aux livres numériques, qu'ils soient les doublons d'ouvrages imprimés déposés ou qu'ils n'existent que sous cette forme, le dépôt légal des livres numériques payants est encore en débat dans les instances parlementaires et en chantier à la BnF. À la différence de celle des sites web, la collecte des livres numériques payants ne peut pas être effectuée automatiquement en ligne. Il a donc été décidé de créer un mode d'entrée spécifique, se rapprochant du modèle du dépôt des ouvrages imprimés.

Si les éditeurs gardent la responsabilité du dépôt légal, la possibilité leur est donnée de mandater leur distributeur en ligne aux fins d'effectuer le dépôt en leur nom. Cette collaboration facilite la levée des mesures techniques de protection ou des formats propriétaires ajoutés par les distributeurs, qui empêchent la gestion et la bonne conservation des livres par la BnF. En outre, afin d'être exploitables par les plateformes de vente en ligne, les livres numériques sont décrits dans des fichiers de métadonnées mises en forme par les distributeurs qui seront récupérées dans le catalogue de la BnF. Après une première expérimentation, une chaîne de traitement complète sera mise en œuvre très prochainement. L'intégralité du processus est conçue pour fonctionner de manière automatisée. Afin de sécuriser l'accès aux collections, auquel les éditeurs sont très attachés, leur consultation se fera via *Gallica intra muros*, dans les salles de recherche uniquement, et les livres ne devront être ni copiables ni imprimables. Les livres numériques apparaîtront néanmoins aussi dans le catalogue général de la BnF.

Il est important désormais de rendre plus facilement visible et consultable,



par les lecteurs et les chercheurs, ce corpus numérique qui est dans le prolongement direct des corpus de livres imprimés : trouver au catalogue le *Tiers Livre* de François Bon au même titre que ses autres livres est devenu nécessaire. La BnF a ainsi déjà entrepris de rendre visible dans son catalogue les notices de quelques centaines de blogs d'écrivains archivés au titre du dépôt légal du web. Il n'est bien entendu pas question de cataloguer le web, mais de proposer une mise en valeur particulière de ce corpus. Il est temps en effet d'agir pour mieux reconnaître, collecter, conserver, et valoriser ces nouvelles écritures littéraires.

Le dépôt légal des écrits numériques soulève on le voit encore beaucoup de questions, mais la réflexion sur la manière d'améliorer les solutions élaborées est déjà bien engagée.

Christine Genin

Christine Genin est chargée de collection en littérature française contemporaine et coordinatrice du dépôt légal du web pour le Département Littérature et Art à la BnF.

Voir aussi : « Le devenir web de la littérature ». *Revue de la BnF*, 52, avril 2016, p. 152-162.