

Livres numériques : plutôt livre ou numérique ? L'extension du dépôt légal aux livres numériques à la Bibliothèque nationale de France

French translation of the original paper: "Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France"

Sophie Derrot

Département du dépôt légal, Bibliothèque nationale de France, Paris, France.
sophie.derrot@bnf.fr

Clément Oury

Département du dépôt légal, Bibliothèque nationale de France, Paris, France.
clement.oury@bnf.fr

Traduction en français par Ange Aniesa

Département du dépôt légal, Bibliothèque nationale de France, Paris, France.
ange.aniesa@bnf.fr



This is a French translation of "Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France" copyright © 2014 by **Ange Aniesa**. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Résumé:

En dépit du lent développement du marché du livre numérique en France, notamment si on le compare avec celui des pays d'Amérique et d'Europe du nord, le nombre de titres accessibles est désormais en expansion. Les chaînes de production et de distribution du livre numérique se sont stabilisées, tandis que des formats standards de livres numériques (EPUB, PDF) et de métadonnées (ONIX) se sont imposés. Un nombre croissant de Français a désormais recours aux livres numériques pour se cultiver et se divertir. Être en capacité de collecter ce nouveau type de documents afin de les préserver à long terme est ainsi devenu un enjeu majeur pour la Bibliothèque nationale de France.

Afin de y répondre et de mettre en place un circuit complet de dépôt automatisé, la BnF bénéficie de son expérience en matière de gestion de document numérique (dans le cadre de sa bibliothèque numérique Gallica) et de son savoir-faire en matière de dépôt légal. Elle compte également s'appuyer sur ses relations établies de longue date avec les éditeurs et les distributeurs de livres numériques. Néanmoins, les bibliothèques doivent composer avec la double nature des livres numériques : à la fois documents numériques collectés dans le cadre du dépôt légal de l'internet, et format inédit de document, qui vient s'ajouter à ceux que la BnF acquiert depuis des siècles.

Cet article a pour but de présenter les problématiques auxquelles la BnF est confrontée du fait de ce nouvel enjeu de collecte et des solutions choisies pour y répondre. Le déploiement technique de ce

circuit complet de dépôt légal – entrée, catalogage, accès, et préservation- vient juste de commencer et ses résultats définitifs sont attendus pour la fin de l'année 2014 ou les premiers mois de 2015.

Mots-clés : préservation, livres numériques, dépôt légal numérique

Un contexte favorable

L'émergence tardive d'un marché du livre numérique en France

À l'orée de l'an 2000, les médias français ont annoncé la naissance d'une publication d'un genre totalement nouveau : le livre numérique ou *ebook*. Des appareils appelés « Softbook », « Cybook » ou encore « Rocket e-books » étaient supposés changer le monde de l'édition et promouvoir un tout nouveau mode de lecture : l'écran allait remplacer le papier. Les entreprises françaises avaient souvent un rôle moteur dans cette révolution auto-proclamée. Néanmoins, parmi les nombreux appareils lancés au début du XXI^e siècle, très peu ont survécu. Trop compliqués, trop chers, ils sont en réalité venus trop tôt et n'étaient pas en mesure de créer un marché du livre numérique en France : le bon vieux livre papier est alors, pour une décennie au moins, considéré comme beaucoup plus fiable que son cousin numérique.

Peu à peu, des innovations technologiques, telle que l'encre électronique, et la popularité grandissante de nouveaux appareils, comme les tablettes, ont fait évoluer cette situation de méfiance. Le marché français est en train de changer. Des éditeurs uniquement présents en ligne font leur apparition. Les plus gros éditeurs commencent à sortir leurs livres directement dans les deux versions. Ils entament la numérisation de leurs collections papier, souvent à l'aide de subventions de l'État. La normalisation devient un vecteur d'innovation technologique. La jungle des formats s'est progressivement simplifiée au profit d'un nombre restreint : le PDF pour les ordinateurs et les tablettes, l'EPUB (et ses équivalents propriétaires tels que KF8) pour les liseuses. En interne, la normalisation pour le partage de métadonnées est en cours, grâce à l'adoption massive du format ONIX¹. Enfin, une normalisation s'opère aussi du côté de la production des livres numériques et de leurs canaux de ventes. De façon schématique, trois principaux acteurs émergent.

Le circuit actuel de production et de distribution du livre numérique

Au début de la chaîne, l'éditeur produit le livre : production intellectuelle, par les auteurs, et production matérielle, sous sa forme papier ou numérique –cette étape peut aussi se voir confiée à un sous-traitant. À l'autre bout, les librairies en ligne sont en charge de la promotion et de la vente du livre numérique au lecteur. Le rôle de l'e-distributeur est de faire le lien entre les deux. Il réceptionne le livre numérique en provenance de l'éditeur, vérifie son format, inspecte et améliore la qualité de ses métadonnées et envoie l'ensemble aux vendeurs en ligne.

En comparaison de ses équivalents internationaux, en particulier aux États-Unis et en Europe du Nord, le marché français du livre numérique n'est pas encore parvenu à maturité. Le déséquilibre entre l'offre papier et l'offre numérique diminue, mais il reste considérable

1. ONIX for Books, une norme en XML élaborée pour favoriser le partage sémantique d'informations. Voir <http://www.editeur.org/83/Overview/>

en terme de demande de la part des lecteurs. Cependant, plusieurs indicateurs montrent que la tendance est à la hausse : en mars 2014, un sondage indique que 15 % des Français lisent des livres numériques, contre 5 % deux ans plus tôt². Être en capacité de collecter ce nouveau type de ressource afin de le préserver sur le long terme est ainsi devenu un enjeu majeur pour la Bibliothèque nationale de France.

La gestion des documents numériques à la BnF : de la numérisation au dépôt légal numérique

La gestion de documents numériques n'est pas chose nouvelle à la BnF. La bibliothèque est impliquée dans la numérisation depuis 1992 avec sa bibliothèque numérique, Gallica. Celle-ci a désormais atteint les deux millions de documents : des livres et des périodiques, des journaux, des gravures, des objets, des enregistrements sonores, des documents audiovisuels...

La BnF a aussi la charge de la réception de documents numériques par le biais du dépôt légal. Le dépôt légal est une très ancienne mission de la bibliothèque, établie en 1537 à la Renaissance par le roi François I^{er}, défenseur des arts et des lettres. Au fil des siècles, la loi a évolué pour couvrir divers genres et formes de publications, s'adaptant de fait à chaque changement technologique et social majeur. C'est particulièrement vrai aux XX^e et XXI^e siècles, lorsque l'innovation en termes de médias a créé de nouvelles formes de documents, qui ont progressivement intégré le champ du dépôt légal. La BnF a reçu ses premières ressources numériques lorsque des documents audio et vidéo, respectivement soumis au dépôt légal depuis 1938 et 1975, furent distribués sous format numérique (en CD puis DVD). En 1992, un prolongement de la loi couvre le cas des logiciels et des bases de données. Le 1^{er} août 2006, l'évolution juridique la plus récente intègre finalement dans le champ du dépôt légal toutes les publications en ligne appartenant au domaine français – la loi étant désormais intégrée au code du Patrimoine [1].

Le dépôt légal de l'internet couvre désormais tous les types de ressources diffusées en ligne : des sites d'actualité aux parties publiques des réseaux sociaux, des portails institutionnels aux blogs. En effet, le dépôt légal n'a pas pour vocation de sélectionner les documents à préserver selon leur valeur intellectuelle ou artistique. Cependant, ce nouveau prolongement provoque deux changements fondamentaux dans les principes du dépôt légal. Tout d'abord, le terme de « dépôt » n'est plus réellement pertinent. La BnF et l'Institut national de l'audiovisuel (INA), pour les sites des télévisions et radios, doivent collecter les documents par des moyens automatisés. Un logiciel de collecte, appelé robot, est utilisé pour récupérer les contenus en ligne. Ensuite, l'exhaustivité, qui est la finalité initiale du dépôt légal, n'est plus un objectif réalisable pour la bibliothèque, confrontée à l'immense masse instable qu'est l'internet. En lieu et place de l'exhaustivité, la BnF poursuit un objectif de représentativité : elle cherche à constituer des collections qui reflètent ce qui est disponible en ligne à une date donnée. La conjugaison d'une collecte large annuelle sur les sites en .fr et de collectes ciblées de sites web sélectionnés par les bibliothécaires de la BnF et d'institutions partenaires, a été la solution retenue pour mener à bien cet objectif [2].

2. Des enquêtes sur les usages du livre numérique sont menées par trois acteurs de la chaîne : la SOFIA, le SNE et le SGDL. Voir http://www.sne.fr/img/pdf/SDL/2012/Barometre_SofiaSneSgdl_Les%20usagesdulivrenumerique_mars2012.pdf et <http://www.sne.fr/img/pdf/Evenements/Assises/Assises-21mars2014/Barometre-SNE-Sofia-SGDL-des-usages-du-livre-numerique-21-03-2014.pdf>.

Puisque la plupart des livres numériques sont proposés et vendus en ligne, ils sont supposés entrer dans les magasins de la BnF par la voie du dépôt légal de l'internet. D'un point de vue juridique, les livres numériques ont donc strictement le même statut que n'importe quelle autre publication en ligne. Néanmoins, d'un point de vue ontologique, ils sont des équivalents numériques d'un type de document que les bibliothèques conservent depuis des siècles. La dualité des livres numériques – à la fois livre et numérique – explique pourquoi la BnF a décidé de leur dédier un mode d'entrée et un canal de traitement spécifiques, qui diffèrent de ceux des archives du web³. Ainsi, la bibliothèque s'appuie sur sa triple expérience de réception des livres imprimés, de numérisation de livres et de collecte de publications en ligne afin d'élaborer un circuit complet de dépôt de livre numérique.

Premières expériences de collecte de documents numériques

Les premières expériences de collecte de documents numériques à la BnF ont commencé vers 2000, avant même que la loi ne change. Ces recherches sont alors menées par le département de la Bibliothèque numérique, qui s'occupe alors également de Gallica. Les deux manières d'obtenir des documents numériques sont étudiées par les bibliothécaires de la BnF : la collecte automatique de contenus en ligne et le dépôt légal directement en provenance des éditeurs de livres et de journaux. Néanmoins, la première approche (la collecte automatisée) est vite considérée comme une priorité et l'effort principal est mis sur la construction d'un circuit complet d'archivage du web. Plusieurs raisons justifient ce choix.

Tout d'abord, les sites web sont alors perçus comme les documents les plus en péril, le web étant un espace très récent de publication : il apparaît crucial d'en préserver les premières formes. Il y avait aussi des raisons plus pragmatiques. D'une part, grâce à la coopération internationale, les outils de l'archivage du web étaient déjà disponibles ; de l'autre, peu de livres numériques étaient disponibles, la chaîne du livre numérique n'était elle-même pas assez mûre et les solutions de dépôt n'étaient pas encore jugées assez stables. Les contraintes étaient également légales : en France, les lois votées par le parlement sont mises en application par décret ; le décret sur le dépôt légal numérique est paru seulement en décembre 2011. Par conséquent, entre 2006 et 2011, il est envisageable de collecter de contenus disponibles en ligne gratuitement (comme c'est le cas de l'immense majorité des sites web), mais pas de demander aux éditeurs de déposer leurs contenus payants, comme les livres numériques. Cela explique pourquoi la BnF n'a pas élaboré de circuit de dépôt légal numérique avant 2012.

L'approche de la BnF en matière de dépôt légal des livres numériques

La coopération, un facteur clé

La coopération avec les éditeurs a immédiatement été identifiée comme un élément crucial pour bâtir un système durable et efficace. La BnF a pu s'appuyer sur sa relation de longue date avec le Syndicat national de l'édition (SNE), principal syndicat français d'éditeurs. En septembre 2012, au cours d'une rencontre avec les représentants du SNE, de la BnF et du ministère de la Culture, fut décidée la création de deux groupes de travail complémentaires :

3. Les livres numériques fournis au public sur support physique (CD, clés USB...) entrent déjà dans les collections de la BnF par le biais du dépôt légal audiovisuel. Mais cela représente une très petite part du marché des livres numériques.

- un groupe de travail juridique, mandaté pour examiner si le décret actuel sur le dépôt légal de l'internet prenait suffisamment en compte les spécificités des livres numériques ;
- un groupe de travail fonctionnel, qui devait poser les principes d'un système de dépôt.

Les travaux du groupe fonctionnel ont débuté en mars 2013. Ses premières décisions ont concerné la manière de récupérer du contenu : un système de dépôt direct a été préféré à la collecte en ligne. La BnF aurait pu choisir de moissonner les plateformes de livres numériques comme elle le fait pour les titres de presse en ligne [3]. Cependant, dans la plupart des cas, les livres numériques ne sont pas directement stockés sur les sites des vendeurs en ligne. Le site du libraire est seulement le lieu où sont effectuées les transactions. Le document en lui-même, ou le lien pour le télécharger, est envoyé à l'acheteur par d'autres biais (par exemple par courriel). Dans ce cas précis, la collecte par le web n'aurait pas été adéquate. En outre, cette solution de collecte de livres numériques n'aurait tiré aucun parti des relations établies entre les éditeurs et les services du dépôt légal imprimé. Ici, les principes en vigueur pour le dépôt légal des imprimés prévalent donc sur la logique de collecte de l'internet : dans le dépôt légal « traditionnel », les libraires ne sont pas en charge du dépôt des documents à la bibliothèque ; ils ne doivent pas non plus l'être pour le dépôt légal numérique.

Le groupe de travail fonctionnel a ainsi préféré laisser la charge du dépôt aux éditeurs. Cependant, pour des raisons pratiques, les éditeurs se sont vus proposer la possibilité de mandater leur distributeur pour le faire. Du point de vue de la BnF, travailler avec les distributeurs est vite apparu comme la meilleure solution :

- Il n'y a qu'un petit nombre de distributeurs, pour des centaines d'éditeurs.
- La BnF bénéficie d'un premier contrôle qualité effectué par les distributeurs, à la fois sur les livres numériques et sur leurs métadonnées.
- Les distributeurs reçoivent les livres numériques sans DRM ; ils sont donc en capacité de les envoyer à la bibliothèque sans DRM (voir plus loin).

Une analyse des chaînes de transmission de documents numériques

Les discussions du groupe de travail furent complétées par plusieurs enquêtes menées par une équipe dédiée à la BnF, regroupant des spécialistes des départements du Dépôt légal et des Systèmes d'information.

Ce groupe est allé à la rencontre des principaux distributeurs français, appelés à devenir les premiers à mettre en place ce dépôt légal. Ils ont étudié les formats de données et de métadonnées utilisés, leur manière de transmettre les contenus et ce qu'ils pouvaient envoyer à la bibliothèque.

Dans le même temps, le groupe a aussi produit une analyse de la chaîne d'entrée des documents numériques de la BnF (des documents issus des collections de la BnF numérisés par des prestataires privés ; des documents audio et vidéo ; des masters de livres pour personnes malvoyantes transmis par des éditeurs à la BnF⁴). Le but de l'enquête était

4. Selon la loi, lorsqu'une association de personnes malvoyantes en fait la demande, la BnF est habilitée à demander aux éditeurs le master numérique d'un livre papier (dans l'idéal une version XML) et de le

d'identifier quels outils et quels chemins d'entrée déjà développés pour des documents similaires pouvaient être réutilisés pour un circuit de dépôt de livres numériques. Ces études ont mené à la constitution d'une première ébauche de circuit à l'été 2013. Des analyses plus poussées ont suivi dans la dernière partie de l'année ; leurs conclusions sont exposées ci-dessous. Les développements concrets ont commencé dans les premiers mois de 2014.

Une mise en œuvre progressive

La définition du périmètre

Comme il l'a déjà été mentionné, le paysage des formats de livres numériques disponibles dans le commerce, qui était plutôt chargé il y a quelques années, a eu tendance à se simplifier au bénéfice de formats standardisés (EPUB) ou des formats propriétaires des grands acteurs du marché. Même si cela ne résout pas tous les problèmes, cette simplification est bienvenue du point de vue du dépôt légal, puisque la loi de 2006 s'applique théoriquement à l'ensemble des déclinaisons des formats produits et commercialisés.

Les efforts se concentrent sur les fichiers qui peuvent facilement être définis comme des livres car partageant les spécificités des livres imprimés. Ainsi, des fichiers en format TXT ou DOC sont exclus dès le départ, car ils sont plus souvent des formats de production que des formats de publication ; ils ne sont pas disponibles dans le commerce et les distributeurs ne les utilisent pas. À l'autre bout de la chaîne, les appareils de lecture et leurs logiciels ne sont pas non plus concernés par le dépôt. Comme expliqué plus haut, les fichiers avec DRM et les formats propriétaires sont aussi à exclure du traitement. Une fois tout ceci filtré, que reste-t-il ? Au cours des débats préalables, les éditeurs et distributeurs ont clairement identifié l'EPUB en version 2 et le PDF comme les formats les plus produits et vendus, le MOBI et l'EPUB en version 3 étant actuellement minoritaires. Le PDF et l'EPUB (versions 2 et 3, à mise en page fixe et reformatable) sont donc les cibles principales.

Premier étape : la réception des livres numériques et de leurs métadonnées

Les distributeurs doivent déposer leurs fichiers dans un fichier compressé ZIP, contenant à la fois les données (l'EPUB ou le PDF) et les métadonnées (un ou plusieurs fichiers ONIX), sur une plateforme FTP dédiée. Les fichiers sont accompagnés de leur empreinte (*checksum*) MD5, pour que la BnF puisse vérifier leur intégrité durant l'ensemble du traitement. Une fois ces fichiers ZIP déposés, une première série de tests doit être exécutée par la bibliothèque : des analyses anti-virus, des vérifications d'emplacement de fichiers, etc. Les fichiers sont ouverts et leurs contenus sont encore de nouveau traités. Avant leur versement dans les collections de la bibliothèque, une période d'attente est définie pour permettre à chaque distributeur d'envoyer des mises à jour bibliographiques et/ou techniques des données et/ou des métadonnées. Une fois écoulé le délai pour ces éventuelles mises à jour, et si l'ensemble fourni a passé toutes les vérifications, le livre numérique reçoit un numéro de dépôt légal et le distributeur est informé que le document a bien été déposé.

Une application supervise le traitement global, de la réception des documents à l'accès et au catalogage spécialisé. Cela permet l'échange d'information entre les différentes applications utilisées et assure la cohérence d'ensemble du traitement.

transmettre à l'association. L'association a ensuite le droit de transformer le livre en une version numérique adaptée. Voir <https://exceptionhandicap.bnf.fr/platon-web/>.

Automatiser la réception des métadonnées et leur conversion

Le principe directeur dans la mise en place du dépôt légal des livres numériques est l'automatisation, incluant la réutilisation des métadonnées créées par les éditeurs et distributeurs pour leur propre besoins. Les formes les plus utilisées pour le transfert de métadonnées dans le marché du livre sont les fichiers ONIX, qui sont généralement créés par le distributeur à partir des informations fournies par l'éditeur et qui sont ensuite transmis aux libraires en ligne avec les fichiers des livres. Les avantages des fichiers ONIX attachés aux livres numériques par les distributeurs sont leur richesse et leur précision : à cause de leur finalité commerciale, les informations fournies se doivent d'être exactes. D'autre part, une petite partie des données sont inutiles pour la bibliothèque (le prix des livres par pays par exemple), tandis que manquent des métadonnées bibliographiques essentielles (l'ISBN de la version imprimée n'est ainsi pas toujours indiqué). Afin de rendre les données ONIX déjà existantes aussi utiles que possible pour la bibliothèque, un modèle de fichier ONIX, intégrant les spécifications propres à la BnF, a été élaboré par les bibliothécaires et proposé aux éditeurs et distributeurs.

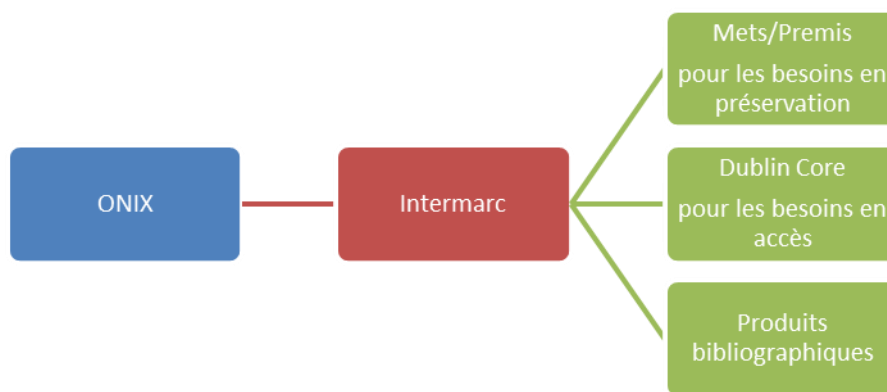


Schéma 1. Une information, des usages multiples.

Dès qu'un livre a passé tous les contrôles de la zone de dépôt, ses métadonnées sont traitées par l'application d'entrée. Cette application convertit notamment des informations issues du fichier ONIX en XML ou INTERMARC, le format spécifique des métadonnées utilisé à la BnF. Ces informations fournies par les distributeurs sont susceptibles être publiées dans le catalogue de la BnF et placées sous licence ouverte.

Cette notice bibliographique est assez riche car les informations en provenance du fichier ONIX sont bien détaillées (titre, auteur, éditeur, mais aussi résumé, notice biographique de l'auteur, mots-clés, etc.). Dans une phase ultérieure du projet, la création automatique de liens va permettre d'établir une relation entre, par exemple, les notices des deux formats d'une même œuvre (papier et EPUB, EPUB et PDF) grâce aux ISBN que les éditeurs vont, a priori, fournir dans le fichier ONIX. Selon le même principe, les identifiants ISNI pourront être utilisés pour relier les notices bibliographiques des livres numériques aux notices d'autorités.

Toutes ces étapes ont vocation à être automatisées autant que possible. Néanmoins, pour certains livres déposés, les notices bibliographiques seront en partie retravaillées par des catalogueurs afin de gagner en précision. Les notices en provenance du catalogue de la BnF sont déjà largement réutilisées par d'autres bibliothèques : les métadonnées des livres

numériques sont plutôt attendues, car il n'existe actuellement pas de bases de données spécifiques pour ce type de métadonnées. Même les éditeurs sont intéressés par cet enrichissement des métadonnées, afin d'améliorer leurs propres bases bibliographiques.

Le flux de données : mettre la préservation au cœur du système

Le livre numérique original doit être conservé. Il ne subira pas de conversion, contrairement aux fichiers de métadonnées. Le dépôt légal est en effet soucieux du format du document et ce format se doit d'être conservé dans son état d'origine.

Comme précédemment exposé, les fichiers de livres numériques passent une première série de tests au moment où ils sont déposés afin de vérifier leur intégrité et format. Un outil de validation développé par l'International Digital Publication Forum (IPDF), EPUBCheck, doit être utilisé pour les fichiers EPUB, mais le choix est toujours en suspens pour les PDF. Le but de ces tests est de vérifier s'il est réellement possible pour la bibliothèque de gérer ces fichiers : les DRM et les formats fermés ne seront pas autorisés par le système, à la fois pour des raisons de préservation et d'accès.

La raison première de l'exclusion des livres numériques avec DRM est que la manipulation des fichiers pendant le processus de dépôt doit être permise. Des copies multiples – parfois bloquées par les DRM – sont de facto nécessaires pour le dépôt, l'accès et le processus de préservation⁵ : mettre en place un système de dépôt légal pour ce type de documents protégés ne semble pas faisable.

Les éditeurs et les distributeurs se sont accordés sur cette position pragmatique. Les formats fermés ne sont pas déposés pour les mêmes raisons : les systèmes de préservation seront probablement incapables de les traiter, notamment sur le long terme, et les solutions d'accès seraient difficiles à mettre en place. En dépit de l'ambiguïté de cette position – en contradiction avec l'objectif traditionnel d'exhaustivité du dépôt légal – il semble plus raisonnable de procéder ainsi afin d'assurer l'accès et la préservation à long terme, qui font également partie des objectifs du dépôt légal.

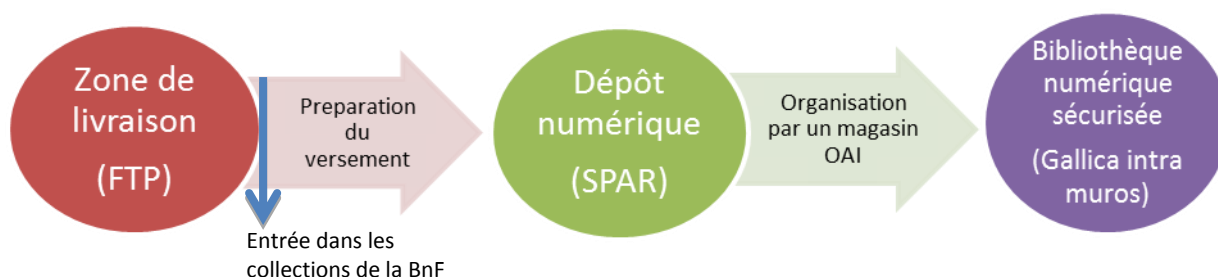


Schéma 2. Le flux de données.

5. Sur les questions de préservation de ressources numériques avec DRM voir [4].

Une fois les livres numériques entrés dans les collections de la BnF, leur intégration à la plateforme de dépôt numérique SPAR (Système de Préservation et d'Archivage Réparti)⁶, est préparée. Une zone tampon dédiée à l'attente de dépôt sera développée pour le circuit du dépôt légal négocié. Le dépôt légal est considéré comme « négocié » car les conditions de dépôt sont discutées entre le distributeur et la bibliothèque, afin par exemple que tous les formats ne soient pris en compte. Ce module va préparer l'ensemble de fichiers à être intégrés, en groupant les fichiers de livres numériques (EPUB et PDF), le fichier ONIX original, si possible un fichier image de la couverture du livre, et enfin le fichier METS⁷. Le rôle de SPAR dans le circuit du dépôt légal est crucial. Tout d'abord, le module de préversement va fournir des informations primordiales pour le catalogue général, tel que l'identifiant ARK du livre numérique, c'est-à-dire son identifiant pérenne. De plus, SPAR va livrer les livres numériques à la plateforme et à l'application d'accès.

Dans SPAR même, deux canaux seront mis en place, en fonction des formats de livres numériques et de leurs caractéristiques :

- Les livres numériques dont le format est considéré comme « géré », après tous les contrôles qualité et de caractérisation seront versés par le premier canal.
- Les fichiers valides, sur le plan du format, mais qui comportent d'éventuels risques en matière de préservation, seront intégrés par le second canal. Par exemple, les fichiers EPUB contenant des éléments en Flash ou Javascript ne sont pas considérés comme « gérés » car la BnF n'a pas assez de garanties sur sa propre capacité à les préserver sur le long terme.

Ainsi, les documents ne seront pas accessibles avant qu'ils ne soient prêts pour la préservation à long terme. Le feu vert sera donné au catalogue et à l'application d'accès uniquement lorsque la plateforme de préservation numérique les aura traités.

L'accès aux documents issus du dépôt légal est juridiquement restreint aux espaces recherche de la bibliothèque et seul un public accrédité peut demander à les consulter. La restriction concerne aussi bien les documents papier que numériques, archives du web et futures livres numériques déposés inclus. Une application spécifique de la bibliothèque numérique, Gallica intra muros, donne accès à du contenu numérique protégé (c'est-à-dire sous droits) et n'est pas disponible hors des murs de la bibliothèque. La sécurité des collections étant un point de vigilance central pour la bibliothèque et ses partenaires, elle a été renforcée pour Gallica intra muros. La solution adoptée devrait être un « navigateur virtuel » qui donne accès aux documents sans aucune possibilité pour l'utilisateur de les télécharger ou de les modifier. Les PDF et les EPUB seront disponibles dans la même application, parmi les autres collections numériques sous droits de la bibliothèque.

L'évolution des formats et des outils de lecture sera un enjeu crucial dans les années à venir. Un dialogue permanent avec les partenaires sera nécessaire pour le suivi de leurs pratiques. En complément, comme la bibliothèque le fait déjà en termes de préservation, la participation de la BnF aux entités internationales comme IDPF pourrait être une bonne solution pour rester au courant des évolutions en matière d'accès.

6. Sur SPAR, voir [5].

7. Le METS (Metadata Encoding and Transmission Standard) est un format XML. Les fichiers METS contiennent toutes les métadonnées nécessaires pour décrire un document numérique et le préserver à long terme.

Quelle place pour l'intervention humaine ?

Il n'est pas encore possible d'évaluer quel degré de travail humain interviendra dans le processus : est-ce qu'un contrôle qualité sera effectué par des bibliothécaires sur les livres numériques avant qu'ils n'intègrent les collections ? Quel pourcentage de livres sera catalogué par les bibliothécaires, si catalogage il y a ? Quel degré de qualité sera jugé acceptable pour les notices de catalogue ?

Dans tous les cas, la gestion du numérique en plus des livres papiers va représenter un bouleversement pour les professionnels du dépôt légal. Afin de les accompagner dans ce changement, le département a créé une série de formations appelées « ateliers numériques ». Tous les bibliothécaires du dépôt légal sont supposés y participer, avec pour grandes problématiques « qu'est-ce qu'un livre numérique ? », « comment est produit, acheté, lu, préservé un tel document ? ». Ce premier volet de formation doit donner à tout un chacun les notions de base et une sensibilisation aux publications électroniques. Une seconde étape serait l'organisation d'exercices concrets de réception, contrôle et catalogage des livres numériques – mais elle interviendra plus tard, lorsque seront effectués les choix cruciaux liés à l'ampleur du travail « manuel » dans le circuit des livres numériques.

Conclusion : les enseignements et les prochaines étapes du projet

Les développements ont déjà commencé et le circuit complet devrait être prêt fin 2014 ou début 2015. Les premières expérimentations avec des distributeurs volontaires vont être menées à la fin de l'année. L'implication des éditeurs et des distributeurs est le point fort du projet. La bibliothèque a la possibilité de travailler avec un nouveau type de professionnels. Certains éditeurs ou distributeurs appartiennent au circuit traditionnel du livre et ont déjà l'habitude de travailler avec la BnF. Mais d'autres partenaires sont relativement nouveaux et spécifiques au marché du livre numérique ; il peut s'agir d'éditeurs uniquement présents en ligne, ayant de nouvelles pratiques et de nouveaux points de vue. La réflexion sur le dépôt légal des livres électroniques doit prendre en considération toute la diversité de sa chaîne de production et de distribution.

Le second point fort du circuit est sa capacité à s'appuyer sur de nombreux autres projets en cours à la bibliothèque. Presque toutes les applications requises existent déjà, sous une forme ou une autre. Comme les équipes concernées travaillent de conserve, la réflexion sur les livres numériques bénéficie de la conjonction d'un grand nombre d'expertise – du marché du livre électronique aux formats, du catalogage aux métadonnées en passant par la préservation, et ainsi de suite. Tout cela devrait assurément favoriser la mise en place du dépôt légal des livres numériques.

Références

[1] Illien G., Sanz P., Sepetjan S., Stirling P. 2012. The state of e-legal deposit in France: looking back at five years of putting new legislation into practice and envisioning the future. In *IFLA journal*, 2012, vol. 38, n° 1. [http://www.ifla.org/files/hq/publications/ifla-journal/ifla-journal-38-1_2012.pdf]

[2] Bonnel, S., Oury C. 2014. La sélection de sites web dans une bibliothèque nationale encyclopédique: une politique documentaire partagée pour le dépôt légal de l'internet à la

BnF. Dans les *Actes de la 80e conférence de l'IFLA* (Lyon, France, août 2014).
[<http://library.ifla.org/998/1/107-bonnel-fr.pdf>].

[3] Oury C. 2011. When press is not printed: the challenge of collecting digital newspapers at the Bibliothèque nationale de France. Dans les *Actes de préconférence de la section presse de l'IFLA* (Mikkeli, Finlande, août 2012).
[<http://www.ifla2012mikkeli.com/getfile.php?file=154> or http://halshs.archives-ouvertes.fr/docs/00/76/90/84/PDF/LegalDepositNewspapersBnF_Oury_IFLA2012.pdf]

[4] APARSEN. 2013. *Report on DRM Preservation*.
[http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2014/02/APARSEN-REP-D31_1-01-1_4.pdf].

[5] Derrot S., Fauduet L., Oury C., and Peyrard S. 2012. Preservation is Knowledge: A community-driven preservation approach. Dans les *Proceedings of the 9th International Conference on Preservation of Digital Objects* (Toronto, Canada, Octobre 2012).
[<https://ipres.ischool.utoronto.ca/sites/ipres.ischool.utoronto.ca/files/iPres%202012%20Conference%20Proceedings%20Final.pdf>]