

Diplôme de conservateur des bibliothèques

Mémoire d'étude / Janvier 2011

Quel dépôt légal pour les *ebooks* ?

Sophie Derrot

Sous la direction de Gildas Illien
Chef du service du dépôt légal numérique - BnF

Remerciements

Je tiens à remercier particulièrement Gildas Illien, qui m'a apporté disponibilité et soutien avec générosité, et qui a mis à ma disposition toute la documentation nécessaire à cette étude.

Je remercie également pour leurs conseils éclairés et précisions indispensables Emmanuelle Bermès, Élodie Bertrand, Anne Bommier-Chasles, Christophe Evans, et Clément Oury.

Merci aussi à Nicolas, Julien et à ma famille pour leur patience.

Résumé :

Les ebooks sont l'un des enjeux futurs du Dépôt légal, et la BnF s'interroge sur le traitement à leur apporter afin de conjuguer l'exhaustivité de sa mission et les difficultés techniques et économiques que recèle la collecte du livre numérique. Cette étude se penche sur deux scénarios de capture, l'un par le biais de l'Extranet éditeurs, l'autre par celui du moissonnage du Web.

Descripteurs :

Dépôt légal

Édition électronique

Métadonnées

Bibliothèque nationale (France)

Abstract:

Ebooks are one of the future stakes of the Legal Deposit. The BnF ponders over their treatment for combining the exhaustiveness of his mission, and the technical and economical difficulties of the ebooks collection. This study examines two possibilities, one by means of the Extranet for editors, and the other by means of the Webharvesting.

Keywords:

Legal deposit

Electronical edition

Metadata

Bibliothèque nationale (France)

Droits d'auteurs



Cette création est mise à disposition selon le Contrat : « **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 2.0 France** » disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/2.0/fr/> ou par courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA.

Sommaire

SIGLES ET ABREVIATIONS	7
INTRODUCTION : LA BIBLIOTHEQUE NATIONALE PASSE AU DEPOT NUMERIQUE	9
CHAPITRE 1 : POUR UNE DEFINITION DE L'EBOOK	11
I. L'ebook, entre support physique et fichier	11
II. D'une définition synthétique à un ensemble de critères.....	13
III. Une question de formats et d'usages ?	14
CHAPITRE 2 : L'EBOOK, UN PRODUIT CULTUREL.....	17
I. L'histoire éditoriale de l'ebook, un scénario en plusieurs actes	17
II. Les acteurs du paysage éditorial français.....	19
III. Un contexte mouvant ?	22
CHAPITRE 3 : LE DEPOT LEGAL.....	25
I. Le contexte du dépôt légal à la Bibliothèque nationale de France.....	25
II. Le dépôt légal numérique et son cadre juridique	26
CHAPITRE 4 : LES PREMICES.....	31
I. Quels formats ?	31
A. Les formats des ebooks	31
B. Les métadonnées.....	32
C. La gestion des droits.....	35
II. Le circuit des données numériques à la BnF	36
III. La mise en place d'une interface de consultation.....	38
CHAPITRE 5 : UN DEPOT PAR L'EXTRANET EDITEURS.....	41
I. L'Extranet éditeurs	41
II. Quelles possibilités pour le Dépôt légal des ebooks ?.....	43
A. Les déposants	43
B. Le dépôt	44
C. Nécessités.....	45
1. L'information en direction des déposants	45
2. Une mise en place phasée	46
3. Coûts et moyens	47
III. Avantages et inconvénients	49
A. Avantages.....	49
B. Inconvénients.....	50
CHAPITRE 6 : UN DEPOT PAR LE MOISSONNAGE DU WEB.....	53
I. Le Dépôt légal du Web	53
II. Quelles possibilités pour le dépôt légal des ebooks ?	54
A. La collecte.....	55
1. Les collectes larges.....	55
2. Les collectes ciblées ou courantes	56
B. Le traitement : référencement, conservation, consultation	57
C. Coûts et moyens.....	60
III. Avantages et inconvénients	61
A. Avantages.....	61

<i>B. Inconvénients</i>	62
CONCLUSION : LE DEPOT LEGAL DES <i>EBOOKS</i>, UNE GAGEURE	63
BIBLIOGRAPHIE	67
Textes de loi.....	70
TABLE DES ANNEXES	71

Sigles et abréviations

AFNIC	Association française pour le nommage Internet en coopération
BnF	Bibliothèque nationale de France
CNIL	Commission nationale de l'informatique et des libertés
CNL	Centre national du livre
DADVSI	(Loi relative) au droit d'auteur et aux droits voisins dans la société de l'information
DAV	Département audiovisuel (BnF)
DCO	Département des collections (BnF)
DDL	Département du dépôt légal (BnF)
DLN	Dépôt légal numérique (service) (BnF)
DSI	Département des systèmes d'information
ETP	Équivalent temps plein
INA	Institut national de l'audiovisuel
SNE	Syndicat national de l'édition

Introduction

Le livre numérique ou *ebook* est encore une relative nouveauté dans le paysage éditorial français. Mais la montée en puissance de ce phénomène depuis quelques années, et particulièrement depuis 2008¹, oblige le monde des bibliothèques à se pencher sur les principes qui régissent l'édition numérique. Si les périodiques électroniques sont désormais bien implantés dans la politique documentaire des bibliothèques françaises, la prise en compte des *ebooks* reste plus diffuse.

La volonté de définir l'objet bibliothéconomique bute dès le début sur la notion même de livre numérique : qu'est-ce exactement qu'un *ebook* ? Le terme de livre entend celui d'une entité intellectuelle, et son opposition au périodique électronique le rapproche de la monographie papier, avec un titre, une table des matières, une cohérence interne. Mais là où l'imagination peut librement s'exprimer, c'est dans les enrichissements divers que l'aspect numérique peut apporter à ce que nous mettons aujourd'hui sur l'idée de livre : les liens hypertextes, les images modulables, le zoom, les vidéos, les annotations, etc. On verra que chercher à fixer une définition rigide de l'*ebook* conduit à un échec, positif pour un objet qui met en avant sa souplesse et sa richesse.

Au vu des évolutions du monde de l'édition vers le numérique, il est indispensable d'envisager une récolte des *ebooks* par le dépôt légal. En effet celui-ci, organisé en France par la BnF, a pour mission de collecter la production éditoriale nationale et diffusée sur le territoire français, dans une volonté de conservation patrimoniale. De fait, jusque récemment, les textes législatifs déterminant le dépôt légal ne prenaient pas en compte les publications par voie électronique. C'est chose faite depuis la loi DADVSI, et la BnF a pu pleinement mettre en place le dépôt légal de l'Internet notamment. La notion de patrimoine touche donc maintenant sans crainte le domaine du numérique, et l'on peut considérer naturellement d'englober les *ebooks* dans la mission patrimoniale du dépôt légal, au même titre que les ouvrages imprimés.

Certes, des questions techniques et organisationnelles se posent sur ce dépôt qui, s'il va intellectuellement de soi, ne peut être comparé à strictement parler avec celui des imprimés. La multiplicité de ses formes de diffusion et de production, les flous qui subsistent sur des points de définition comme sur des points techniques et législatifs brident encore la mise en place d'un « DLE » (dépôt légal des *ebooks*). Ne serait-ce que par leur place entre un dépôt légal du Web en plein essor et un dépôt des imprimés qui se modernise, les livres numériques sont au cœur des problématiques du DDL.

La rapide obsolescence des formats et des accompagnements technologiques des *ebooks* oblige à se pencher rapidement sur la question de leur capture et de leur conservation sur le temps long qui est celui du dépôt légal. Le contexte privilégié de la diffusion des livres numériques, Internet, évolue plus rapidement que les autres environnements éditoriaux que traite la BnF. L'exigence de flexibilité et d'ouverture donne au sujet du livre numérique un dynamisme exaltant, sur un domaine encore en construction.

Cette étude s'appuie sur des réflexions préalables au sein de la BnF et du monde de l'édition, car les deux secteurs s'interrogent de manière parallèle sur cette même question, sans toujours se rencontrer. Sur le terrain du dépôt légal pourtant, un dialogue étroit doit être entretenu entre éditeurs-producteurs et l'établissement dépositaire. Sera d'abord évoqué le contexte, à la fois intellectuel, éditorial et institutionnel du livre

1. « 2008 : Les ebooks sont partout. » : M. LEBERT. *Une courte histoire de l'ebook*. NEF, Université de Toronto, 2009. Disponible à l'adresse : <<http://www.etudes-francaises.net/dossiers/ebook.htm>>.

numérique et du dépôt légal. Ensuite, deux propositions de méthode de récolte seront présentées, en tenant compte de structures qui existent déjà, l'Extranet éditeurs et le dépôt légal du Web.

Chapitre 1 : Pour une définition de l'*ebook*

Les termes d'*ebook* ou de « livre numérique » recouvrent une réalité qui ne fait pas l'unanimité et qui se compose d'éléments variables. Pour la suite de notre travail, nous devons reprendre les termes et leurs usages. Le manque d'harmonisation qui existe en ce domaine entre les différents acteurs peut cependant empêcher de s'arrêter sur une définition stricte, du fait des évolutions technologiques de ce marché.

I. L'*EBOOK*, ENTRE SUPPORT PHYSIQUE ET FICHER

L'*ebook*, souvent indistinctement traduit par livre numérique ou livre électronique, a historiquement d'abord été considéré comme dépendant totalement d'un support de lecture. C'est sans doute ce qui justifie que pour certains, ce terme désigne avant tout l'objet physique, le dispositif électronique permettant la consultation du fichier numérique : « c'est un outil dédié et il se distingue en cela de l'ordinateur classique [...]. La différence vient de la mobilité, du poids (généralement inférieur à 300 g), de l'absence de clavier physique et de sa spécialisation d'usage². » En 2005, la commission générale de terminologie et de néologie retient ainsi comme définition de l'*ebook* celle d'un « support électronique en forme de tablette comportant une mémoire vive et permettant la lecture d'un texte sur écran³. » L'Éducation nationale reprend exactement ce principe. Cet attachement à un dispositif fait indéniablement partie de la notion de livre numérique, ne serait-ce que pour servir de support à la représentation du livre numérique comme alternative au livre papier. Cependant, l'y résumer peut paraître réducteur.

Parallèlement à cette dimension de support de lecture, le terme d'*ebook* peut également désigner le fichier ou l'ensemble de fichiers numériques en lui-même, qui transpose le document papier sous une forme numérique. Ainsi Bernard Prost dans son rapport sur l'édition numérique⁴ donne clairement l'*ebook* comme « tout ouvrage publié dans un format électronique [...], il s'agit d'un fichier ou un ensemble de fichiers », alors que le nom de « *reader* » reste au dispositif dédié de lecture ; il va jusqu'à désigner le livre papier comme « *p-book* » (p pour « *print* »), en opposition à l'*ebook*. Cette définition est acceptée par d'autres, dont Bruno Patino dans son rapport sur le livre numérique⁵. Par ailleurs, il faut noter que le grand public aurait tendance à rejoindre ces vues, puisque pour plus de la moitié des Français interrogés en 2009, le livre numérique est un livre sous forme de fichier numérique, alors qu'un tiers seulement pense plutôt à une tablette électronique dédiée à la lecture⁶.

2. B. PROST. *Rapport d'étude sur l'édition numérique de livres scientifiques et techniques, l'éditeur des années 2010* : étude réalisée par QUAE avec le soutien du ministère de l'Enseignement supérieur et de la Recherche. 2008. p. 5. Disponible à l'adresse : <http://media.education.gouv.fr/file/2008/41/8/rapporteditionnumerique_28418.pdf>

3. Définition du livre électronique de la commission générale de terminologie et de néologie, publiée au *JO* du 18 janvier 2005.

4. B. PROST, *op. cit.*

5. B. PATINO. *Le devenir numérique de l'édition : du livre objet au livre droit* : rapport au MCC. La Documentation française, 2008. Série des rapports officiels. Disponible à l'adresse : <<http://lesrapports.ladocumentationfrancaise.fr/BRP/084000381/0000.pdf>>

6. Centre national du livre, IPSOS MediaCT, *Publics du livre numérique* : enquête menée de septembre 2009 à février 2009. Disponible à l'adresse : <<http://www.centrenationaldulivre.fr/?Le-livre-sera-t-il-numerique/>>

La presse a d'abord adopté la définition liée au support⁷, mais il semble que récemment, la généralisation du phénomène ait modulé sa position, et qu'elle désigne peu à peu le dispositif physique sous le nom de « *reader* », « liseuse » ou de « liseuse électronique », voire tout simplement « tablette »⁸. Le manque d'harmonie générale est cependant particulièrement visible dans le domaine journalistique. Outre-Manche, la confusion semble être similaire à tous les niveaux, et la volonté d'une définition harmonisée tout aussi présente, comme le montre l'article de Magda Vassiliou et Jennifer Rowley, « Progressing the definition of "e-book"⁹ ». Ces dernières favorisent d'ailleurs la solution qui assimile l'*ebook* à l'objet numérique.

Le terme *ebook* recouvre donc bien deux réalités : l'objet électronique, la « liseuse », et l'ouvrage lui-même sous forme de fichier(s). L'absence de consensus sur la définition de l'*ebook* est certes préjudiciable au dialogue entre les acteurs, mais permet de se poser les bonnes questions. Comme le remarque Nathalie Borg¹⁰, il faut désormais réaliser par exemple que contrairement à la situation du livre papier, le contenu et le contenant sont deux choses bien distinctes avec le livre numérique, que les barrières avec certaines autres formes de communication (vidéo, musique, etc.) doivent être sérieusement repensées, et la simple offre de livres homothétiques dépassée.

De ce constat d'une nécessité d'ouverture sont issues des tentatives de définitions très larges du livre, incluant numérique comme imprimé. Celle proposée le 13 janvier 2009 par le syndicat national des éditeurs tient à resituer l'idée de livre dans la publication électronique :

« Un livre est un ensemble de données textuelles, graphiques ou sonores résultant d'un travail éditorial, publié sous un titre à une date déterminée et ayant pour objet la reproduction d'une œuvre de l'esprit d'un ou plusieurs auteurs en vue de l'enseignement, de la diffusion de la pensée et de la culture. Cet ensemble destiné à être lu peut être présenté sous la forme d'éléments assemblés ou réunis par tout procédé d'impression ou numérique éventuellement interactif, sous réserve que ces éléments aient le même objet et que leur réunion soit nécessaire à l'unité de l'œuvre. Ils ne peuvent faire l'objet d'une vente séparée que s'ils sont destinés à former un ensemble ou s'ils en constituent une mise à jour. Cet ensemble conserve la nature de livre lorsque la surface cumulée des espaces consacrés à la publicité et des blancs intégrés au texte en vue de l'utilisation par le lecteur est au plus égale au tiers de la surface totale de l'ensemble, abstraction faite de la reliure ou de tout procédé équivalent¹¹. »

Dans la même idée de globalité, le Parlement européen définit comme un livre : « toutes les œuvres imprimées ou reproduites sous une autre forme, en particulier de la littérature, la musique, l'art et la photographie, les magazines spécialisés (à l'exception

7. Voir par exemple : <<http://www.usinenouvelle.com/article/livre-electronique-enfin-mur-pour-le-decollage.147206>>, ou : <http://www.lexpress.fr/actualite/high-tech/le-livre-numerique-pour-les-nuls_473207.html> (consultés le 19 novembre 2010).

8. Voir par exemple l'article « De nouvelles liseuses électroniques arrivent sur le marché français », dans *lemonde.fr*, paru le 22 octobre 2010, disponible à l'adresse : <http://www.lemonde.fr/technologies/article/2010/10/22/de-nouvelles-liseuses-electroniques-arrivent-sur-le-marche-francais_1430059_651865.html> (consulté le 19 novembre 2010). À noter que dans cet article, l'auteur utilise le terme « *ebook* » pour l'offre américaine, et celui de « liseuse » pour l'offre française. On a particulièrement pu observer l'utilisation de ce vocabulaire lors du récent lancement du FnacBook.

9. M. VASSILIOU, J. ROWLEY. Progressing the definition of "e-book". *Library Hi-Tech*, 2008, vol. 26, n° 3. p. 355-368. Disponible à l'adresse : <<http://www.emeraldinsight.com/journals.htm?issn=0737-8831&volume=26&issue=3&PHPSESSID=0shpla9aqbrikqr377q9ad6p2>>.

10. N. BORG. *Régime juridique du livre : de l'imprimé au numérique* : mémoire de master 2 Nouvelles technologies de l'information et de la communication, université de Saint-Quentin-en-Yvelines. Florence Gaullier (dir.). Septembre 2010. Disponible à l'adresse : <<http://www.juriscom.net/uni/visu.php?ID=1255>>.

11. Cité dans N. BORG, *op. cit.*, p. 7.

toutefois des quotidiens et hebdomadaires ou des revues à grande diffusion) et les éditions électroniques, dans la mesure où elles remplacent des livres imprimés¹². »

On distingue clairement la difficulté de l'étude et du traitement bibliothéconomique de l'*ebook* en dressant ce premier panel de définitions, qui ne couvrent pourtant que le support sur lequel le livre est présenté, sans entrer dans le détail de son contenu ou de la forme de celui-ci.

II. D'UNE DEFINITION SYNTHETIQUE A UN ENSEMBLE DE CRITERES

Au sein-même de la BnF, la difficulté à assurer des contours clairs à l'objet *ebook* est perceptible. Pourtant, cette définition a des enjeux majeurs pour le traitement de ce qui doit maintenant être considéré par la bibliothèque comme une réalité. Sur demande du Comité de coordination du dépôt légal, le service du dépôt légal numérique, secondé par le service juridique, s'est penché sur cette question, qui a fait l'objet de deux notes, le 20 juillet 2009 et le 17 mars 2010.

Constatant les difficultés pour le moment assez insolubles que nous avons soulignées, la proposition est faite de les contourner et d'élaborer une liste de critères discriminants qui formeraient les caractéristiques premières d'un *ebook*. Ce procédé répond bien sûr aux besoins de la bibliothèque en termes de dépôt légal, mais elle permet aussi de rendre plus concret l'*ebook* comme voisin et parallèle numérique au livre traditionnel. Les critères qui découlent de cette opération indispensable d'identification et de différenciation dessinent les usages que le lecteur et le bibliothécaire pourront faire de l'*ebook*, qui après tout un objet assez neuf et dont l'utilisation n'est pas encore entrée dans nos habitudes.

Le DLN propose tout d'abord, en relation avec ce que nous avons vu plus haut, de considérer le livre numérique comme un contenu, c'est-à-dire un fichier, et non un contenant, un support de lecture. Ce point permet d'envisager plus largement le livre numérique comme non pas rattaché uniquement à un objet physique : en effet, dans l'état actuel du droit touchant au dépôt légal¹³, les outils de lecture ne font pas l'objet d'un dépôt, et s'ils sont conservés par la bibliothèque, c'est avant tout à des fins pratiques (lecture des formats rattachés à un support) ou à titre de *specimen*. De plus, cette acception induit un traitement plus complet et sans doute plus souple et plus pérenne, qui dépendrait moins des évolutions technologiques et des variations de supports (liseuse, mais aussi téléphone, console de jeux, etc.). Dans l'*ebook*, le texte est donc présenté indépendamment de son support.

Intéressant les repères traditionnels du livre, il est recommandé d'envisager l'*ebook* comme devant présenter les « caractères internes » d'une monographie contemporaine ; l'*ebook* doit simplement ressembler à un livre imprimé. Ici on retrouve ce qui dans notre esprit participe de la représentation du livre : une page de titre, une pagination continue, un sommaire, une table des matières. Quelles que soient les possibilités actuelles ou futures du livre numérique en terme de navigation interne par exemple, ces éléments participent à la fois des repères au sein du livre et de sa structure. Liaison est ainsi faite entre la vision traditionnelle d'un ouvrage et la modernité de sa version numérique, ce qui revêt notamment une certaine importance pour les éditeurs, et un côté pratique pour le bibliothécaire et le lecteur.

12. Résolution du Parlement européen du 16 mai 2002 contenant des recommandations à la Commission en vue de l'élaboration d'une directive du Parlement européen et du Conseil de l'Union européenne relative à un système de prix imposés pour les livres.

13. Voir chapitre 3.

Il importe cependant de faire ici une distinction : notre travail porte sur le livre numérique, qui se différencie du livre numérisé¹⁴, également présent dans les préoccupations de la BnF (Gallica) et du paysage éditorial. Il s'agit d'une publication faite pour le numérique, qu'elle soit une nouvelle publication d'un ouvrage existant déjà sur papier, ou bien qu'elle soit « née numérique ». Nous laissons également de côté le e-journal, version numérique du périodique ; l'*ebook* est ici spécifiquement considéré hors de toute parution périodique.

III. UNE QUESTION DE FORMATS ET D'USAGES ?

Si l'on s'arrête à considérer l'*ebook* comme étant un fichier ou ensemble de fichiers, il est impératif de se pencher sur les formats de ces fichiers. En effet, ce sont de ces derniers que dépendront à la fois les possibilités offertes par les *ebooks*, mais également les contraintes qui guideront les choix des bibliothèques dans le traitement de ces documents (mise à disposition, conservation). La question des formats participe en partie de la difficulté à considérer le paysage actuel des *ebooks*. On pourrait envisager une définition de l'objet par ces formats (« Un *ebook* est un fichier se présentant sous tel format »), mais ceux-ci recouvrent encore une réalité assez contrastée et trop variée pour que ce soit réellement possible.

Quels sont les formats les plus utilisés ? Le site Numilog met à disposition dans sa rubrique d'aide un tableau recensant les principaux formats actuellement disponibles¹⁵.

Parmi ces formats¹⁶, on trouve principalement :

- le Portable Document Format (PDF), créé par Adobe Systems : très utilisé, il n'est pas conçu à l'origine pour l'*ebook*, mais sa spécificité est de préserver la mise en forme du document définie par l'auteur, quelle que soit l'application utilisée pour la lecture du fichier ; ses principaux inconvénients restent son adaptation encore difficile aux différentes tailles d'écran, et son manque de flexibilité et d'adaptation aux technologies intuitives ;
- le format EPUB, un format libre et standard recommandé par l'IDPF¹⁷ : basé comme beaucoup de formats d'*ebooks* sur le langage XML, il permet notamment une pagination dynamique ; il bénéficie d'une grande portabilité et a le grand avantage d'être lisible sur la plupart des supports et des applications ;
- le Mobipocket PRC, destiné à être lu avec le logiciel gratuit Mobipocket Reader¹⁸ qui peut être installé sur Palm Pilot, Pocket PC, PC, smartphone, ou le Kindle d'Amazon ;
- le format BBeB (LRF), spécifique au Sony Reader et propriétaire¹⁹ ;

14. Cette différenciation a notamment été relevée lors de la table ronde *Stratégies éditoriales à l'heure du livre électronique* du 9 février 2009, organisée par la bibliothèque Landowski de Boulogne-Billancourt. Voir un compte rendu de cette rencontre sur le blog du BBF, à l'adresse : <<http://bbf.enssib.fr/blog/2009/02/25/le-livre-numerique-table-ronde-du-9-fevrier-2009-a-boulogne-billancourt>> (consulté le 19 novembre 2010).

15. Ce tableau, reproduit dans son état du 20 novembre 2010, est consultable en annexe 1, et à l'adresse <<http://www.numilog.fr/aide.asp>>.

16. On trouvera une liste complète et des tableaux comparatifs détaillés sur la page anglaise de Wikipédia dédiée à la comparaison entre les formats d'*ebooks* : <http://en.wikipedia.org/wiki/Comparison_of_e-book_formats> (consulté le 21 novembre 2010).

17. L'International Digital Publishing Forum (IDPF) compte notamment des éditeurs de logiciels (Adobe), des libraires et des éditeurs (Hachette, JSTOR ou Harlequin), mais aussi Google ou Apple.

18. Certaines plateformes de téléchargement et de vente d'*ebooks* disposent d'une rubrique permettant le téléchargement d'outils de lecture, au nombre desquels le Mobipocket Reader, ou bien le logiciel Adobe Digital Editions (ADE).

19. Cependant Sony a annoncé en juillet 2010 que le format BBeB serait progressivement abandonné au profit de l'EPUB pour la lecture des *ebooks* sur les Sony Reader.

- le format Microsoft Reader (LIT), destiné à être lu avec le logiciel de Microsoft, uniquement sur ordinateur (Pocket PC ou PC) ;
- le format Palm Doc, qui est issu du format PDB, celui des bases de données contenues dans le PalmPilot, et qui est aussi lisible sur un grand nombre de smartphones.

Les formats texte, DOC et ODT voire TXT, sont plus problématiques, puisqu'ils ne sont pas réservés à l'édition d'ebooks : ce sont davantage des formats de création que de sortie, comme ceux cités précédemment ; ils peuvent donc difficilement être pris comme formats déterminants. Cependant, certaines plateformes, comme Ebookslibresetgratuits.com, mettent à disposition les sources de leurs ouvrages libres de droits dans des documents texte zippés, afin que l'utilisateur puisse les convertir au format qui lui convient le mieux. Il faut remarquer que la plupart des distributeurs d'ebooks vendent également des livres audio, au même titre que les ouvrages sur les formats nommés ci-dessus. Ces livres audio sont dans les formats numériques MP3 en majorité, ou WMA (Microsoft).

Le moyen le plus répandu pour contrôler l'utilisation faite des œuvres numériques reste l'application à ces formats de dispositifs de gestion des droits numériques (en anglais *Digital Rights Management*, DRM). Il s'agit d'empêcher une utilisation et surtout une diffusion non contrôlées. Ainsi, le logiciel ADE conditionne la lecture à l'inscription d'un mot de passe ; de même, l'iPad d'Apple lit les ebooks achetés sur l'iBookstore au format EPUB, mais en y accolant ses propres DRM, mettant en place un système similaire à celui du Kindle d'Amazon. Si ces DRM sont utilisées sous l'argument de défense des droits des auteurs, il faut avoir conscience qu'elles auront sans aucun doute de lourdes conséquences sur la pérennité de la lecture des œuvres ainsi protégées²⁰. De plus, ces modes d'exploitation verticaux, construits sur le modèle mis en place par Apple pour la musique, font courir un risque de cloisonnement du marché, et rendent d'autant plus complexes les réflexions sur ce sujet pour les bibliothèques.

Le format a donc son importance dans le processus de mise à disposition : certains formats sont propriétaires, d'autres ouverts, certains ne permettent la lecture que sur un seul support dédié, d'autres sont plus souples. Cette question de la portabilité, c'est-à-dire la capacité d'un format à être lu sur différents supports, est centrale, l'interopérabilité totale étant encore loin d'être atteinte. Aucun format ne permet une portabilité complète, à la fois sur ordinateur, quelque soit le système d'exploitation (PC, Linux, Mac), sur téléphone portable ou sur liseuse. On peut cependant noter que des logiciels sont développés petit à petit pour permettre la lecture de certains formats sur différents supports non prévus à l'origine (pour les systèmes Linux, pour l'iPhone d'Apple, etc.), et que l'EPUB présente sans aucun doute le plus de souplesse, en s'adaptant à quasiment tous les supports.

Il faut remarquer que la plupart des livres vendus avec DRM dans ces formats interdisent l'impression. Seul le PDF peut l'autoriser, selon la volonté de l'éditeur.

Comme nous allons le voir dans le chapitre suivant, certains secteurs de l'édition sont plus aptes à basculer dans le numérique. Ces tendances dépendent en partie des usages qui sont rendus possibles par l'ebook, selon son format. La duplication à l'identique du « PDF-imprimeur » n'est plus d'actualité²¹ et il s'agit d'offrir au texte une réelle plus-value en l'enrichissant notamment par les possibilités multimédias²². Comme

20. Sur le sujet de la gestion des droits numériques, voir également le chapitre 4, I.C.

21. H. BIENVAULT. *Le coût d'un livre numérique : étude réalisée pour le MOTif*. Avril 2010. Disponible à l'adresse : <http://www.lemotif.fr/fichier/motif_fichier/153/fichier_fichier_etude_sur_le.coa.t.d.un.livre.numerique.pdf>.

22. Voir à ce sujet la rencontre-débat sur le livre « enrichi », le 29 novembre 2010 à la BPI, disponible en podcast dans les archives sonores de la bibliothèque (<http://archives-sonores.bpi.fr/index.php?urlaction=doc&id_doc=3327>, consulté le 9 mars 2011).

le relèvent M. Vassiliou et J. Rowley²³, les *ebooks* présentent des caractéristiques typiques, qui renforcent l'interactivité et la navigation interne : la possibilité d'effectuer des recherches dans le texte, des liens hypertextes vers le Web²⁴, des fonctions d'annotation et de surlignage, des objets multimédias et outils interactifs, la possibilité de changer la taille de la police, de zoomer. Tous les formats ne permettent pas ces usages, et il y a fort à parier que ce sont ces derniers qui feront le tri dans le choix actuel. De plus, l'*ebook* prend également place au sein d'un environnement qui ajoute encore des fonctionnalités externes, comme un catalogue en ligne, des plateformes qui permettent à la fois la vente et/ou le téléchargement, mais également le commentaire et le conseil d'ouvrages, la fourniture d'outils de lecture, le rapprochement d'autres ouvrages, imprimés comme numériques. L'*ebook* se place au centre d'une constellation d'usages, selon les fonctionnalités qu'il offre, ce qui le rend novateur²⁵ et intéressant, mais d'autant complexe à envisager du point de vue bibliothéconomique.

Dans l'optique du dépôt légal, les exigences de conservation et de consultation influent sur la préférence que l'on peut accorder à certains formats. C'est de la portabilité que dépend une grande partie des problématiques autour de la conservation et de la mise à disposition du fichier. Ainsi, on préférera les formats ouverts, voire normalisés (PDF, EPUB). Au contraire, les formats issus de modèles de distribution verticaux rendent plus difficiles le traitement des fichiers (Apple, Sony ou Amazon). Le format EPUB pourrait être un bon terrain d'accord, même s'il subit quelques critiques pour les livres ayant une mise en page avancée, comme les bandes dessinées ou les livres techniques. Cependant, sa bonne portabilité joue pour lui. Il faut également prendre en compte les réflexions autour de l'environnement de l'*ebook*, selon les usages qu'il permet : que conserver parmi les services qui l'entourent ? comment le faire ?

« Définir un livre numérique relève de la gageure²⁶ » et cette définition dépendra très certainement de l'évolution future du marché éditorial en ligne et des choix qui seront fait dans ce domaine.

23. M. VASSILIOU, J. ROWLEY, *op. cit.*

24. Les liseuses qui sont commercialisées permettent de plus en plus la connexion à Internet, par la 3G, le Wifi ou Bluetooth.

25. Ces usages ont été lancés pour la plupart par la librairie Amazon, d'abord avec son offre papier, puis étendue à l'offre numérique et répandue sur d'autres plateformes.

26. B. PATINO, *op. cit.*, p. 56.

Chapitre 2 : L'ebook, un produit culturel

Pour remettre en contexte la situation de l'ebook comme produit culturel, nous proposons d'en retracer rapidement l'histoire. Il s'agit de percevoir le caractère encore mouvant et incertain sur quelques points du paysage éditorial qui se développe autour de cette technologie. Bien avoir en tête les avancées – et les reculs – qui ponctuent l'histoire de l'ebook permet également de prendre toutes les questions posées par son traitement avec prudence.

I. L'HISTOIRE EDITORIALE DE L'EBOOK, UN SCENARIO EN PLUSIEURS ACTES

Marie Lebert, dans sa *Courte histoire de l'ebook*²⁷, prend comme point de départ de l'aventure la date de juillet 1971, avec la mise en place aux États-Unis du Projet Gutenberg, comme la première bibliothèque numérique. Pour le fondateur, Michael Hart, il s'agissait alors de diffuser gratuitement par voie électronique le plus grand nombre d'œuvres littéraires possible. Les textes sont numérisés au format ASCII par des volontaires, d'abord essentiellement en anglais, puis, au fur et à mesure que le projet s'étend, à d'autres langues. Au tout début de l'entreprise, le web n'existe pas encore, mais ce projet ouvre la voie à la numérisation et à la mise à disposition gratuite d'œuvres du domaine public. Avec la généralisation du web à partir du début des années 1990, le nombre d'œuvres augmente et le public s'étend. Aujourd'hui, le projet continue et se diversifie à d'autres documents que les textes (livres audio, partitions musicales), bien que ceux-ci restent au centre de sa démarche. Son site web²⁸ indique la mise à disposition de plus de 33.000 ebooks gratuits, libres de droit aux États-Unis, disponibles sous plusieurs formats (EPUB, Kindle, texte brut, HTML, MP3, etc.) et accessibles sur la plupart des supports de lecture.

Le Projet Gutenberg marque le début de l'ebook et de sa propagation gratuite, et concerne surtout le monde anglo-saxon. Le domaine éditorial anglophone connaît une nette avance dans le numérique par rapport au domaine francophone. Ainsi, le premier éditeur à utiliser des ebooks disponibles gratuitement en ligne est un éditeur universitaire américain, la National Academy Press ; et surtout, la première librairie en ligne, Amazon.com, voit le jour en juillet 1995 à Seattle, et met en place sa librairie numérique dès 2000. En France cependant, l'ABU²⁹, qui propose en ligne gratuitement des œuvres libres de droits, est mise en place dès 1993.

Les premiers pas de l'édition numérique commerciale se font à partir du milieu des années 1990. Comme le relève Marie Lebert, cette édition bénéficie d'avantages certains sur l'édition traditionnelle : pas de stock, des coûts de fonctionnement moins élevés, une diffusion plus facile. Elle entraîne aussi une remise en cause d'une part du système car les éditeurs traditionnels et les libraires ne voient pas toujours d'un œil positif l'arrivée de cette nouvelle technologie qui complexifie le système de l'édition et de la distribution :

27. M. LEBERT. *Une courte histoire de l'ebook*. NEF, Université de Toronto, 2009. Disponible à l'adresse : <<http://www.etudes-francaises.net/dossiers/ebook.htm>>.

28. <<http://www.gutenberg.org/>>

29. L'Association des Bibliophiles Universels, hébergée par le conservatoire national des arts et métiers de Paris (CNAM), <<http://abu.cnam.fr/>>.

« On voit des éditeurs traditionnels vendre directement leurs titres en ligne, des éditeurs électroniques commercialiser les versions numérisées de livres publiés par des éditeurs traditionnels, des libraires numériques vendre les versions numérisées de livres publiés par des éditeurs partenaires, sans parler des auteurs qui choisissent de s'auto-éditer sur le web ou de promouvoir eux-mêmes leurs œuvres publiées, et des nouvelles plateformes d'édition littéraire qui se chargent de découvrir de nouveaux talents pour pallier les carences de l'édition traditionnelle³⁰. »

En 1996, Olivier Gainon fonde CyLibris, maison basée à Paris, pionnière dans le monde de l'édition numérique commerciale. Les auteurs des textes publiés par CyLibris sont nouveaux ou confirmés, et les ouvrages ne sont vendus que sur le web, imprimés à la commande et envoyés directement au client : pas de stock, pas d'intermédiaire, mais toujours un aboutissement papier. À partir de 2001, les sites Numilog et Mobipocket proposent les versions numériques de certains titres de CyLibris, que la Fnac vend également en version papier. Deux ans après CyLibris, une autre maison d'édition fait son apparition dans le domaine du livre numérique, 00h00, qui se propose non de vendre des livres imprimés, mais directement des *ebooks* via Internet (au format PDF)³¹. Parmi ces ouvrages, on compte à la fois des œuvres originales, nées numériques, et des rééditions d'ouvrages publiés par d'autres éditeurs. Parallèlement à ces entreprises éditoriales, la fin des années 1990 a vu la commercialisation de ces premiers supports de lecture dédiés, le SoftBook Reader et le Rocket eBook, mis sur le marché américain en 1998, ou le Cybook, projet français à la même date.

Mais ces prémices subissent un échec : les premiers modèles de liseuses sont trop lourds, trop peu indépendants en terme d'énergie, trop complexes à utiliser, trop peu communicants³², et se soldèrent par des faillites, comme celle de la société française Cytale en 2002³³. Parallèlement, en 2000, 00h00 est rachetée par la société américaine Gemstar, également détentrice des entreprises ayant lancé les premières tablettes de lecture, avant cesser son activité en 2003. Les bouleversements qu'introduit le numérique dans l'industrie de la musique semblent donner raison aux plus réticents et aux plus sceptiques et l'édition semble laisser à ce moment l'épisode du livre numérique de côté.

Malgré tout, les réflexions sont lancées et après ce coup d'arrêt, se mettent en place peu à peu les structures qui vont participer à l'essor du livre numérique tel qu'on peut l'envisager aujourd'hui. En France, on pense au lancement de Gallica en 1997, ou bien à la mise en ligne en 2000 du site Numilog, la première librairie francophone à vendre exclusivement des livres numériques (voir plus bas). Les entreprises ont investi dans les recherches sur l'encre et le papier numériques, afin de rendre les liseuses plus pratiques et plus agréables. Ces dernières années, les exemples de la vidéo et de la musique appellent tout à la fois la prudence et l'émulation. L'arrivée dans le secteur du livre numérique de grandes sociétés de l'informatique sonne sans doute la prise d'assurance du développement du livre numérique : Apple (iPad), Amazon (Kindle), Sony (Sony Reader), Microsoft (Microsoft Reader), Google eBookstore. De surcroît, comme le rappelle Stéphanie Chevrier³⁴, il faut rappeler que le monde de l'édition

30. M. LEBERT, *op. cit.*, p. 31.

31. L'éditeur propose également, en complément, l'impression à la demande, mais celle-ci ne représente en 2000 que 15 % des ventes, selon M. Lebert (*op. cit.*, p. 33).

32. C. KREBS. Du livre électronique à l'encre électronique. Paris : BBF, 2006. *BBF*, vol. 51, n° 4. p. 38-41. Disponible à l'adresse : <<http://bbf.enssib.fr/consulter/bbf-2006-04-0038-006>> (consulté le 21 novembre 2010).

33. S. CHEVRIER. Plaidoyer pour le livre numérique. *Laviedesidees.fr* [en ligne], 6 octobre 2008. Disponible à l'adresse : <<http://www.laviedesidees.fr/Plaidoyer-pour-le-livre-numerique.html>> (consulté le 21 novembre 2010).

34. S. CHEVRIER, *ibid.*

travaille désormais en numérique, qu'il s'agisse des auteurs, des éditeurs, des maquettistes, des fabricants ou des imprimeurs : seul l'objet final est imprimé sur du papier.

Par ailleurs, ces développements technologiques génèrent des interrogations dans la société et posent le livre numérique comme un enjeu économique et culturel. Depuis quelques années, plusieurs études et rapports officiels sont parus sur la question du numérique et de la numérisation, dont certains s'interrogent plus particulièrement sur l'édition française face à la montée du phénomène (ceux de Bruno Patino³⁵ et de Bernard Prost³⁶), sur les valeurs que doit véhiculer le marché du livre numérique (celui de Christine Albanel³⁷), ou bien enquêtent sur les publics de l'ebook³⁸.

Cependant, fait révélateur à la fois de la nouveauté du phénomène et de la difficulté encore ressentie à le cerner, il reste encore assez difficile de trouver des chiffres similaires à ceux qui existent pour l'imprimé, tels que ceux de la production d'ebooks par an et par secteur, ou le nombre de livres nés numériques. Le seul document récent que nous ayons pu trouver est une étude du MOTif de mai-juin 2010³⁹ dressant un état des lieux de 31 plateformes de diffusion, et estimant l'offre dans le domaine des ebooks entre 60 et 70.000 titres (commercialisés et sous droits). Cette offre est donc relativement faible, estimée à 10 % de l'ensemble des titres disponibles à la vente papier ; cela dit, ces catalogues évoluent très rapidement. En revanche, le nombre d'inédits numériques reste inconnu.

II. LES ACTEURS DU PAYSAGE EDITORIAL FRANÇAIS

Après les évolutions des dernières décennies, on trouve aujourd'hui un paysage éditorial français du livre numérique assez varié, même s'il reste encore à l'aube de son développement. Les principaux distributeurs d'ebooks sur l'Internet francophone proposent au téléchargement des ouvrages gratuits ou payants, sous des formats divers, destinés aux différents supports de lecture proposés sur le marché, avec parfois la proposition de logiciels de lecture. Certains sont évidemment plus importants que d'autres, mais on n'observe aujourd'hui aucune domination écrasante du marché par l'un ou quelques-uns de ces acteurs. Le monde de l'édition française se laisse doucement pénétrer par le phénomène des ebooks et le rôle des e-distributeurs est central dans la diffusion de ceux-ci.

Ces libraires numériques négocient avec les éditeurs le droit de vendre les ebooks sur une plateforme sécurisée, selon deux modèles : soit la consultation en ligne avec un abonnement, soit l'achat à l'exemplaire avec téléchargement du fichier acquis⁴⁰. Pour les offres disponibles par l'intermédiaire d'une liseuse (Kindle, iPad, Fnacbook), l'utilisateur a la possibilité de se connecter à une librairie numérique qui lui permet de télécharger directement sur son appareil les ouvrages qu'il achète sur la plateforme dédiée, après

35. B. PATINO, *op. cit.*

36. B. PROST, *op. cit.*

37. C. ALBANDEL, et autres. *Pour un livre numérique créateur de valeurs* : rapport remis au Premier ministre. 2009. Disponible à l'adresse <<http://www.ladocumentationfrancaise.fr/rapports-publics/104000189>>.

38. Enquête CNL-IPSOS, *op. cit.*

39. M. DAVAL, R. DOUINE. *Ebookz, l'offre légale et illégale de livres numériques*. Paris : MOTif, mai-juin 2010. 33 p. Disponible à l'adresse : <http://www.lemotif.fr/fichier/motif_fichier/171/fichier_fichier_.l.offre.de.livres.numa.riques.en.france.le.motif.pdf> (consulté le 2 janvier 2010).

40. I. ANTONUTTI. Les ressources électroniques généralistes. *Les collections électroniques, une nouvelle politique documentaire*. P. CARBONE, F. CAVALIER (dir.). Paris : Electre-Le Cercle de la librairie, 2009. p. 19.

une inscription préalable. Les e-distributeurs offrent également pour la plupart des services d'édition.

Que ce soit sur un support dédié ou bien sur un site de distributeur⁴¹, l'acquisition se fait comme dans une librairie classique : l'internaute navigue dans le site par « rayons »⁴², soit thématiques, soit par format, soit par prix, et fait son choix d'*ebooks* à mettre dans son « panier ». Pour le paiement et les services additionnels sur un site, une identification est la plupart du temps nécessaire, à l'aide d'une adresse de courriel et d'un mot de passe, obtenu par une inscription sur le site. Après le choix éventuel du format et l'achat, le téléchargement de l'ouvrage se fait directement par un lien envoyé au client par courriel lui permettant de télécharger un fichier à lire dans le logiciel de lecture de son choix. En ce qui concerne l'offre payante⁴³, cette identification par une inscription est la règle⁴⁴, et permet notamment le paiement sécurisé et les services qui entourent le strict achat de l'*ebook* (bibliothèque numérique, historique des achats, etc.). L'*ebook* est ainsi accessible soit directement depuis le Web, soit sur des plateformes de téléchargement qui ne sont pas liées au site de l'éditeur.

On peut faire une liste des libraires numériques dominantes en France :

- Numilog⁴⁵ : la plus ancienne librairie numérique, rachetée par Hachette en mai 2008. Sa page d'accueil dénombre plus de 50.000 titres disponibles⁴⁶, qui sont proposés selon leurs formats (PDF, EPUB, PRC, MP3 ou WAV), avec des extraits gratuits (PDF). Sont présentés des livres gratuits, une vente au chapitre de certains des ouvrages, des réductions sur le prix éditeur pour d'autres. Ces ouvrages proviennent des éditions du groupe Hachette mais ne s'y limitent pas. La librairie propose un service d'autopublication⁴⁷.
- Cyberlibris⁴⁸ : cette librairie propose un accès forfaitaire depuis 2001, et s'adresse de manière originale à des segments de clientèle (institutions académiques⁴⁹, entreprises, familles) en leur proposant des bibliothèques numériques, et en mettant en place des communautés d'échanges autour de ces livres. Chaque ouvrage est également accessible à l'achat individuel en ligne. 1.500 titres payants au catalogue, qui comprend des livres récents, anciens ou épuisés.
- Ebookslibresetgratuits.com⁵⁰ : ce site francophone ouvert édite et publie des *ebooks* à partir de textes libres de droits, scannés puis OCRisés. Plus de 2.000 ouvrages sont disponibles⁵¹, dans plusieurs formats (PRC, Sony Reader, EPUB, PDF, sources parfois proposées en format texte zippé). Le site met à disposition de ses usagers un catalogue

41. Voir la comparaison des deux procédures d'achat (Fnac.com et IBookStore) dans l'étude de M. DAVAL et R. DOUINE. *Ebookz*, *op. cit.*

42. Le terme est notamment utilisé par Numilog.

43. Ainsi, sur Ebookslibresetgratuits.com, les liens de téléchargement sont disponibles directement sur le site, tout comme ce qui entoure l'ouvrage (par exemple les commentaires des lecteurs ou la possibilité d'en laisser un).

44. Nous verrons dans le chapitre suivant que cette barrière du mot de passe conditionne en partie notre sujet.

45. <<http://www.numilog.fr/>>

46. 52.332 au 28 novembre 2010.

47. <<http://www.jepublic.com/>>

48. <<http://www.cyberlibris.com/>>

49. Le site <<http://www.scholarvox.com/>> s'adresse ainsi tout particulièrement aux écoles de commerce ou d'ingénieurs, avec un choix d'ouvrages en économie, gestion, management et sciences.

50. <<http://www.ebooksg gratuits.com/>>

51. 2.156 au 28 novembre 2010.

OPDS⁵² de son contenu, et permet le téléchargement d'outils de lecture, de création et de conversion d'*ebooks*.

- Eyrolles⁵³ : cet éditeur-libraire propose essentiellement des livres pratiques et professionnels, aux formats EPUB et PDF surtout, ainsi que PRC et zippés. Un flux RSS est disponible pour les nouveautés. Une dizaine d'*ebooks* gratuits est proposée, ainsi que des extraits (PDF).
- Mobipocket⁵⁴ : cette plateforme a été rachetée par Amazon en 2005. Il s'agit de l'offre la plus complète d'*ebooks* actuellement, avec plus de 120.000 titres disponibles selon la page d'accueil du site, dont 10.000 gratuits. La plateforme permet également le téléchargement des logiciels de lecture et de publication d'*ebooks* de Mobipocket.
- Publie.net⁵⁵, « le contemporain s'écrit numérique » : lancé le 1^{er} janvier 2008, cet éditeur-distributeur propose près de 360⁵⁶ textes contemporains nés numériques, mis en page pour la lecture sur support numérique, sans DRM, imprimables. Les formats disponibles sont : PDF, PRC, EPUB, liseuse en ligne (streaming), MP3 ; mais il existe également un accès par abonnement, avec ou sans téléchargement. Le catalogue de Publie.net est également disponible sur l'iBookstore d'Apple.
- Éditions Harlequin⁵⁷ : cette maison d'édition fonctionne en partenariat avec Numilog pour ses *ebooks*, et propose des livres aux formats PDF, EPUB, WMA et MP3.
- Fnac.com⁵⁸ : ce distributeur s'est mis au livre numérique il y a peu. Il propose des ouvrages aux formats EPUB et/ou PDF, avec des DRM, mais également des « Classiques gratuits ». La librairie numérique de la Fnac est disponible directement sur le Fnacbook, et offre, d'après le site, 80.000 *ebooks* en téléchargement.
- Dialogues⁵⁹ : site de la librairie brestoise, qui propose la VPC d'ouvrages papier et la vente directe d'*ebooks* ; bien que ce site ait pris des positions contre les DRM, les conditions de vente précisent que celles-ci peuvent verrouiller les *ebooks* proposés. Les livres sont vendus à l'acte, en PDF ou en EPUB, parmi un catalogue de plus de 10.000 titres numériques.

Quelques-un(e)s de ces plateformes ou éditeurs collaborent actuellement avec Gallica : Numilog, Cyberlibris, mais aussi La Documentation française, Dalloz, Gallimard ou CAIRN. Gallica met à disposition ses premiers *ebooks* gratuits, à partir d'ouvrages libres de droits (EPUB, PRC, HTML ou TXT) ; mais à côté de ces collections patrimoniales, elle expérimente un accès à des documents contemporains, sous droits, grâce à des partenariats avec ces e-distributeurs : quelques extraits sont proposés en feuilletage, et le lecteur est renvoyé au site distributeur s'il désire acquérir l'ouvrage.

L'offre éditoriale est donc assez variée, bien que le nombre d'interlocuteurs soit (encore ?) limité. Les e-distributeurs ont une place centrale dans le système, qu'ils

52. *Open Publication Distribution System*, élaboré par la fondation Internet Archive, est un format de syndication pour la publication électronique, permettant de suivre les nouveautés d'un catalogue d'*ebooks*, mais également de les obtenir.

53. <<http://izibook.eyrolles.com/>>

54. <<http://www.mobipocket.com/en/homepage/default.asp?language=fr>>

55. <<http://www.publie.net/>>

56. Chiffre au 28 novembre 2010.

57. <<http://numerique.harlequin.fr/>>

58. <<http://www.fnac.com/telecharger-ebook.asp>>

59. <<http://www.librairiedialogues.fr/>>

soient éditeurs ou seulement vendeurs. Mais les éditeurs traditionnels ont peut-être un rôle plus important à jouer dans le futur, avec la numérisation de leurs fonds. Cette numérisation a déjà commencé chez certains (Gallimard), et elle est encouragée par l'État qui, par l'intermédiaire du CNL, propose des aides aux éditeurs dans cette optique. Les éditeurs, qui disposeraient alors non seulement d'un fonds de livres numériques contemporains, mais aussi déjà publiés, resteront-ils dépendants des e-distributeurs ou profiteront-ils de leur propre site pour mettre eux-mêmes leurs ouvrages à la disposition du public ? Aujourd'hui cependant, on n'observe pas encore de raz-de-marée du né numérique : les catalogues des éditeurs comptent en majorité des titres qui existent déjà sur papier, soit parce qu'ils sont anciens, soit parce qu'il s'agit de nouveautés également disponibles en imprimé.

III. UN CONTEXTE MOUVANT ?

Il est important de garder à l'esprit que, si certaines tournures sont prises, les changements restent tout à fait possibles dans le domaine de l'édition, particulièrement en ce qui concerne l'avenir de l'*ebook* : la question des formats, mais aussi celle des supports et celle des distributeurs sont autant d'éléments qu'il convient encore de traiter avec une certaine vigilance.

Des tendances peuvent néanmoins être dégagées, au moins en ce qui a trait au paysage actuel. Ainsi, des secteurs précis de l'édition sont d'ores et déjà davantage concernés par l'éclosion d'un marché de l'*ebook*, pour des raisons pratiques, comme le relève Bruno Patino : ils sont confrontés à « des incitations à la dématérialisation car leur présence laisse deviner que l'ouvrage numérique apporte une solution plus satisfaisante que le papier aux besoins des lecteurs⁶⁰ ». En effet, certains types de contenus s'adaptent particulièrement bien au numérique⁶¹ : l'édition professionnelle (sciences, techniques, médecine, droit), les dictionnaires et les encyclopédies (qui avaient déjà fait un pas hors du papier avec les versions sur cédérom), les livres pratiques (guides de voyage, livres de cuisine) ; on peut également penser aux livres scolaires⁶² et aux ouvrages universitaires. Le livre numérique offre pour certaines de ces publications des avantages indéniables par rapport au papier :

- un faible coût de publication, même si cette question fait débat ;
- des possibilités accrues d'indexation, essentielles pour les dictionnaires, les encyclopédies et les livres de voyage ;
- une mise à jour régulière possible à bas coût et rapidement (ouvrages de droit, guides de voyage) ;
- un poids sans commune mesure avec les versions papier (encyclopédies) et donc des facilités de transport (guides, manuels) ;
- une navigation interne étendue entre les différentes parties de l'ouvrage ;

60. B. PATINO, *op. cit.*, p. 18.

61. On trouvera en annexe 2 une évaluation, selon la mesure du possible, des différents secteurs sur quelques plateformes de diffusion (Numilog, Eyrolles, Mobipocket). Les secteurs qui ressortent sont les sciences humaines (dont l'histoire et la philosophie), le monde de l'entreprise (management), les ouvrages de vie pratique et technique, et certains secteurs de la littérature (romans, poésie).

62. Voir les expérimentations actuelles sur les cartables numériques. Par ailleurs, de plus en plus de manuels proposent déjà un contenu enrichi à partir de l'ouvrage papier, accessible par Internet selon diverses procédures (code barre, code, flashcode, etc.).

- des possibilités de renvoi vers d'autres ressources numériques (collaboration récente entre Larousse et l'Encyclopedia Britannica ; lien entre un guide et des sites d'hôtellerie.

Ces avantages touchent également la littérature générale, même s'ils sont là bien moins déterminants. Mais il semble que la littérature soit davantage concernée par un modèle hybride de couplage imprimé/numérique, l'ouvrage paraissant de manière quasi simultanée sur les deux supports, ou bien le numérique suivant la vente du papier. La numérisation des fonds d'éditeurs, ainsi que la création d'*ebooks* à partir de classiques libres de droits fait pour le moment pencher la balance. Cependant, il arrive que la chronologie soit inversée entre les deux supports⁶³, et il faut noter l'arrivée encore timide d'ouvrages nés numériques et vendus exclusivement en tant que tels, comme ceux que propose Publie.net. Dans le domaine de la littérature, tout reste également à voir en fonction de l'évolution des usages, comme a pu le montrer la rencontre-débat organisée par la BPI du 29 novembre 2010⁶⁴.

La facilité de publication explique en grande partie le succès de l'auto-édition sur Internet ; on a vu que la plupart des e-distributeurs/éditeurs offraient de tels services, en réduisant au maximum les frais de publication, et des sites se sont spécialisés dans ce domaine. Dans ce cas, soit l'ouvrage est imprimé à la demande, soit il est directement commercialisé sous la forme d'*ebook* sur la plateforme en question. Dans le contexte du livre numérique, l'auto-édition n'est pas limitée aux œuvres qui auraient du mal à rentrer dans le circuit traditionnel de l'édition ; certes, celles-ci forment une grande partie des *ebooks* auto-édités, mais on peut également trouver des auteurs plus reconnus qui préfèrent se détacher d'une maison d'édition et directement mettre en œuvre la publication de leurs œuvres⁶⁵.

En revanche, certains autres domaines de l'édition se montrent plus réticents au passage à l'*ebook*. On a vu avec la question des formats que la bande dessinée par exemple trouve encore des difficultés à être convaincante sur support numérique, même si le succès des « blogs BD » pourrait participer à une certaine redéfinition du genre⁶⁶. Par ailleurs, les éditeurs commencent à réfléchir à l'exploitation numérique de la bande dessinée, ne serait-ce que pour contrer un piratage grandissant en ce domaine⁶⁷ ; on pourrait notamment regarder du côté du Japon, encore une fois, et des mangas diffusés sur téléphone portable. Cette résistance se constate également dans le secteur des beaux livres ou de la littérature jeunesse⁶⁸.

Il importe de connaître ces spécificités sectorielles de l'édition française pour notre sujet. Elles permettent de mieux repérer les acteurs auxquels il faudra s'adresser et les caractéristiques des ouvrages à recueillir.

63. L'exemple le plus cité est le succès au Japon des romans pour téléphone portable, et le tirage sur papier du plus connu d'entre eux, *Koizora* (2005), vendu à 1,5 million d'exemplaires.

64. À cette occasion ont été exposés plusieurs ouvrages de littérature au contenu enrichi, pour des vidéos, de la musique ou des lectures audio. Ces démonstrations ont pu donner un aperçu des possibilités qu'offre le livre numérique à l'écrit littéraire, si celui s'affranchit du cadre de l'ouvrage tel qu'on le connaît aujourd'hui sous sa forme papier, non interactive, non multimédia.

65. Dans la sphère anglophone, voir l'exemple récent de J.A. Konrath et de son *Draculas*, sorti en auto-édition, dans l'article où il explique ses raisonnements et sa démarche : « eBooks and the ease of self-publishing », paru le 16 octobre 2010 sur le site de *The Huffington Post*, disponible à l'adresse : <http://www.huffingtonpost.com/ja-konrath/ebooks-and-self-publishing_b_764516.html> (consulté le 27 novembre 2010).

66. Le site Ave!Comics propose également les « lectures animées » des bandes dessinées qu'il met en vente (<<http://www.avecomics.com/#/accueil/>>).

67. Voir le paragraphe que consacre à la bande dessinée Hervé Gaymard dans son rapport (p. 103) : *Situation du livre : évaluation de la loi relative au prix du livre et questions prospectives* : rapport au MCC. Mars 2009. Disponible à l'adresse : <<http://www.centrenationaldulivre.fr/?Situation-du-livre-Evaluation-de>>.

68. Conscients du marché, les e-distributeurs mettent pourtant en avant les productions numériques jeunesse sur leurs sites.

Chapitre 3 : Le dépôt légal

I. LE CONTEXTE DU DEPOT LEGAL A LA BIBLIOTHEQUE NATIONALE DE FRANCE

Nous nous contenterons ici d'un rapide rappel du cadre du dépôt légal à la BnF afin de poser notre réflexion⁶⁹.

Institué par François I^{er} en 1537 pour les imprimés et étendu petit à petit jusqu'à l'Internet français en 2006, le dépôt légal a pour but de rassembler la mémoire du patrimoine culturel diffusé sur le territoire français. En organisant la collecte, la conservation et la consultation de documents de toute nature, il a peu à peu permis la constitution d'une collection de référence sur la production éditoriale française. Il est également chargé de l'élaboration et de la diffusion de bibliographies nationales (livres, publications en série, musique, audiovisuel, cartographie).

Le dépôt légal est régi par le Code du patrimoine pour les aspects législatifs, et par le décret n° 93-1429 du 31 décembre 1993 et des arrêtés de 1995 et 2006 pour les aspects réglementaires. Par ces textes, sont concernés par l'obligation de dépôt⁷⁰ :

- « les documents imprimés, graphiques, photographiques, sonores, audiovisuels, multimédias, quel que soit leur procédé technique de production, d'édition ou de diffusion, [...], dès lors qu'ils sont mis à la disposition d'un public » ;
- « Les logiciels et les bases de données [...] dès lors qu'ils sont mis à disposition d'un public par la diffusion d'un support matériel, quelle que soit la nature de ce support » ;
- « [...] les signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique. »

Cette obligation gratuite s'impose à tout producteur des documents concernés, qu'il soit éditeur, imprimeur, distributeur, importateur. Selon le type de dépôt, la BnF partage son statut d'organisme dépositaire avec l'Institut national de l'audiovisuel, le Centre national du cinéma et de l'image animée, et les bibliothèques habilitées par arrêté du ministre chargé de la culture. En ce qui nous intéresse, il est important de relever que le dépôt légal concerne avant tout le support mis à disposition du public, et non l'œuvre, qui peut donc être déposée sous différents supports s'il y a lieu.

Le département du dépôt légal (DDL) est responsable de la collecte et du traitement bibliographique des imprimés concernés par cette obligation de dépôt — les documents spécialisés étant quant à eux traités par les départements spécialisés (Audiovisuel, Estampes, Musique, Cartes et plans) —, ainsi que des documents numériques en ligne. Le DDL prend donc en charge la gestion de la collecte, dans un objectif de continuité des collections et d'exhaustivité (parfois relative comme nous le verrons plus bas). La conservation, la valorisation et la communication des documents incombent aux départements des collections, le DDL gère donc avant tout des flux, depuis la réception jusqu'au catalogage et à la bibliographie, en passant par le service des échanges.

69. Pour des développements plus poussés, on pourra se référer aux pages dédiées sur le site de la bibliothèque (<http://www.bnf.fr/fr/professionnels/depot_legal.html>) ainsi qu'à l'ouvrage de Daniel Bécourt, *Dépôt légal : de l'écrit à l'électronique* (Paris : Litec, 2001, 200 p.), ou au mémoire de DCB de Olivier Delangle, *Le département du dépôt légal à la Bibliothèque nationale de France* (2003).

70. Code du patrimoine, livre I^{er}, titre III, ch. 1^{er}, art. L131-2, modifié par l'ordonnance n° 2009-901 du 24 juillet 2009, art. 5.

Ce département est organisé en 6 services, qui se calquent sur le parcours et le type du document : 154 personnes au 31 décembre 2009 (146,8 ETP) sont réparties entre les services de la gestion des livres (26 agents) et de celle des périodiques (33), les services de la Bibliographie française – Livres (64) et – Périodiques (9), le service des échanges (17), et celui du dépôt légal numérique (5).

Le versement des documents s'accompagne pour les déposants d'une déclaration de dépôt⁷¹ qui permet d'accélérer le traitement et de recueillir les informations minimales avant le catalogage. Afin de faciliter le dépôt des imprimés, le DDL a mis en place depuis le 2 juillet 2009 la possibilité pour les éditeurs de faire leur déclaration de dépôt sur un portail extranet prévu à cet effet, appelé « Extranet éditeurs », sur lequel nous reviendrons (chapitre 5). Grâce à cette interface, les éditeurs déposants inscrits au service (environ 20 %) peuvent également suivre le traitement de cette déclaration par les services du dépôt légal, et accéder aux notices des livres déposés. Il faut cependant relever que quelques éditeurs importants ne semblent pas intéressés pour le moment par ce processus, comme Gallimard⁷², mais qu'il recueille un certain succès auprès des petits éditeurs⁷³.

II. LE DEPOT LEGAL NUMERIQUE ET SON CADRE JURIDIQUE

Afin de s'adapter aux nouveaux supports de diffusion et d'expression de l'information et du savoir, le dépôt légal concerne également depuis 2006 les signes, signaux, écrits, images, sons ou messages de toute nature faisant l'objet d'une communication au public par voie électronique⁷⁴, c'est-à-dire par le biais d'Internet.

La mission patrimoniale du dépôt légal touche donc désormais aussi tous les types de publications diffusés sur Internet, qui sont considérés comme faisant partie intégrante de la production éditoriale diffusée sur le territoire français. Le point de vue adopté dans ce domaine est très large et concerne aussi bien les sites commerciaux, institutionnels ou personnels que les revues, les blogs, les plateformes de vidéos ou, ce qui nous intéresse au premier chef, les livres numériques. En sont exclus la correspondance privée et les intranets. L'objectif d'exhaustivité énoncé plus haut comme étant une règle du dépôt légal connaît ici une certaine relativisation, ne serait-ce que pour des raisons économiques et techniques ; sont donc concernés par ce dépôt numérique les contenus édités ou les éditeurs ayant un lien avec le territoire français⁷⁵ : l'ensemble des sites Web enregistrés en .fr et en .re, grâce à une liste obtenue par une collaboration avec l'AFNIC⁷⁶, ainsi que d'autres sites (.com, .org, .net, etc.) édités par des personnes physiques ou morales domiciliées en France. La bibliothèque procède par échantillonnage, et ne peut recueillir par exemple toutes les mises à jour de tous les

71. On trouvera en annexe 3 deux exemples de cette déclaration de dépôt, l'un pour les documents informatiques (logiciels et bases de données), et l'autre pour les documents multimédias qui associent sur un même support deux ou plusieurs documents.

72. À l'exception notable de son département Jeunesse, qui utilise fréquemment l'Extranet éditeurs.

73. 82 % des éditeurs effectuent entre 1 et 10 dépôts par an, selon les chiffres du dépôt légal (rapport SIPIL-DL de mars 2010), ce qui fait des petits éditeurs la majorité des déposants en général, et 46 % des utilisateurs de l'Extranet sont des petites structures (employant moins de 5 permanents, mais distinctes de l'auto-édition).

74. Art. L131-2 du Code du patrimoine, modifié par l'ordonnance n° 2009-901 du 24 juillet 2009, art. 5.

75. Les sites relevant de la communication audiovisuelle (radio, télévision, etc.) sont quant à eux déposés à l'INA.

76. L'association française pour le nommage Internet en coopération (AFNIC) est l'autorité nationale gérant les noms de domaines en .fr (France), .re (Réunion) et .tf (Terres australes et antarctiques françaises). On compte au mois de septembre 2010 presque 1,8 million de domaines en .fr (AFNIC-Télécom SudParis, *Observatoire 2010 du marché des noms de domaine en France*, éd. 2010, disponible à l'adresse <<http://www.afnic.fr/>>).

documents, même sur un Internet français ainsi délimité. Il est nécessaire d'accepter dans ce domaine un caractère partiel et lacunaire de la collection, qui perd en exhaustivité ce qu'elle gagne en représentativité. La vérification de l'appartenance des sites moissonnés au champ d'application du dépôt se fait a posteriori ; aucune exclusion de principe n'est faite sur la base du contenu des sites (pornographie, publicité, etc.).

Contrairement au dépôt légal traditionnel, l'éditeur ou le producteur ne prend pas ici une part active au dépôt de sa production. C'est la BnF qui s'occupe de la collecte numérique de manière automatique, à l'aide de robots moissonneurs ; elle ne prend contact avec l'éditeur qu'en cas d'inaccessibilité d'une partie du site, pour des raisons techniques (base de données, formulaire, protection par mot de passe) ou commerciales (contenu payant, abonnement). Ce type de dépôt a été confié à un service particulier, intégré au DDL, et travaillant avec des correspondants répartis sur toute la BnF.

Juridiquement, ce dépôt légal numérique prend place dans le cadre du Code du patrimoine, dans la partie relative au dépôt légal, mais aussi dans un régime d'exception au droit d'auteur, élargi par la loi 2006-961 du 1^{er} août 2006 relative au droit d'auteur et aux droits voisins dans la société de l'information (DADVSI), dans son titre IV relatif au dépôt légal. La BnF est autorisée⁷⁷ à procéder à la collecte par des procédures automatiques ou par accord avec le producteur, et la mise en œuvre d'un code ou d'une restriction d'accès ne peut faire obstacle à la collecte. Les éditeurs et producteurs ne peuvent interdire ni la collecte, ni la mise en consultation du dépôt quand celle-ci se fait sur place, par des chercheurs accrédités, sur des postes individuels à usage réservé⁷⁸, comme c'est le cas au niveau Recherche de la BnF.

De plus, la BnF a le droit⁷⁹ de reproduire une œuvre sur tout support et par tout procédé technique, en vue de sa collecte, de sa conservation et de sa consultation sur place, sans avoir à recourir à l'autorisation explicite de l'auteur de chaque œuvre ni à verser de rémunération. Ce dernier point est crucial, non seulement en ce qui concerne les *ebooks* mais également tous les fichiers informatiques, car il permet les processus de conversion et de copie nécessaires à la conservation et à la consultation pérenne.

Si les documents multimédias ou les bases de données ont une place dans ces textes relatifs au dépôt légal, aucune notion de « document électronique » n'y est spécifiée, et on ne peut donc considérer ce type de documents comme une catégorie soumise au dépôt. Toutefois, il existe notamment un dépôt légal de substitution, qui permet à la BnF de demander aux déposants un fichier numérique à la place de l'ouvrage imprimé, selon des accords propres à chaque déposant⁸⁰. C'est le cas aujourd'hui pour les affiches de grand format, qui sont déposées au département des Estampes sous forme numérique. Il est intéressant de considérer ce mode de dépôt, dit aussi dépôt négocié, et qui bénéficie d'une filière de conservation particulière, car il s'agit du dépôt d'un document qui n'existe pas sous cette forme numérique autrement que pour le dépôt légal. Cet aspect législatif offre donc des possibilités intéressantes pour le dépôt de livres numériques.

Les textes de loi abordent donc tous les points nécessaires : les catégories de documents concernés, la collecte, même avec l'obstacle du mot de passe, la consultation, la conservation. Cependant, après consultation de la CNIL, le Conseil

77. Art. L131-2-1 du Code du patrimoine, modifié par la loi DADVSI, art. 41.

78. Art. L132-4 du Code du patrimoine, modifié par la loi DADVSI, art. 42.

79. Code du patrimoine, *ibid.*

80. Décret n° 93-1429 du 31 décembre 1993 relatif au dépôt légal, modifié par le décret n° 2006-696 du 13 juin 2006 : « Par dérogation au premier alinéa, la Bibliothèque nationale de France peut demander le dépôt d'un fichier numérique se substituant au dépôt du document imprimé, graphique ou photographique. Les modalités de ce dépôt sont définies en accord avec les déposants. »

d'État doit encore publier un décret d'application de la loi DADVSI qui devra fixer les conditions précises de sélection et de consultation des informations collectées. Ce décret n'étant pas encore publié, les opérations que mènent la BnF et son DLN se basent actuellement sur des interprétations de la loi, sans avoir les moyens d'aller jusqu'au bout du raisonnement et de la collecte⁸¹. Ainsi, sans ce décret précisant les applications effectives de la loi, ce que l'on vient d'énoncer reste délicat à mettre en œuvre : par exemple, l'étendue de la notion de territoire national applicable à Internet n'est pas explicitée ; il faudrait également spécifier les modalités de la collecte et la répartition des domaines entre l'INA et la BnF.

Pour le sujet spécifique des *ebooks*, l'absence du décret entrave la collecte des publications payantes, en rendant difficile la demande aux éditeurs de sites des mots de passe et identifiants nécessaires, qui selon la loi, ne peuvent pourtant faire obstacle à la collecte, mais qui sont très répandus, comme on a pu le voir. Une expérience de collecte a été réalisée en mars 2009⁸² sur les *ebooks* gratuits disponibles sur quelques sites de e-distributeurs, afin de considérer les possibilités techniques d'une telle opération. Capturer des contenus gratuits limite le risque juridique que pose une telle opération en l'état actuel des choses⁸³, et il faut noter que, pratiquement, le DLN récolte des *ebooks* dans ce contexte depuis plus de cinq ans, au fil de ses moissonnages du Web.

Pour résumer, et en l'état actuel de la loi, et dans l'attente du décret d'application, la BnF peut collecter au titre du dépôt légal : les *ebooks* gratuits diffusés en ligne⁸⁴, les contenus payants étant en attente ; les *ebooks* sur support, considérés comme des bases de données⁸⁵ ou des documents multimédias⁸⁶ ; et les *ebooks* qui se limitent à un fichier texte diffusé sur un support⁸⁷. Actuellement, les *ebooks* diffusés hors Web mais sans support ne sont donc pas pris en compte.

Nous reviendrons dans le chapitre 6 sur les possibilités qu'offre la collecte d'Internet pour les *ebooks*.

Ces premiers chapitres nous ont permis de poser une définition du périmètre de notre objet d'étude, et de son contexte. L'*ebook* que nous considérons répond à une description aussi délicate que complexe, ce qui peut paraître normal de la part d'une réalité qui tend et va de plus en plus tendre à modifier notre vision du livre. Le dépôt légal se doit de prendre en compte cette réalité, afin notamment de ne pas passer à côté de la capture des *ebooks* « incunables », les premiers nés numériques qui commencent à fleurir sur Internet. Il s'agit donc de réfléchir au plus vite à des procédures de dépôt pour ces ouvrages. Si le décret d'application de la loi de 2006 doit être finalisé et publié pour

81. Pour un exemple de réflexion autour de ce décret, voir l'intervention de Valérie GAME, « Le dépôt légal : dernières évolutions législatives et réglementaires », lors des journées d'information et d'échange du Réseau des bibliothèques de dépôt légal imprimeur à la BnF, les 22 et 23 janvier 2009, p. 5, disponible à l'adresse <http://www.bnf.fr/documents/dl_loi.pdf>.

82. Concernant cette expérience, voir le chapitre 6.

83. Des accords spécifiques ont d'ores et déjà été mis en place avec certains éditeurs pour la fourniture de ces mots de passe, comme le portail de revues CAIRN.

84. Au titre du dépôt légal de l'Internet, art. L131-2 du Code du patrimoine.

85. Au titre du dépôt légal des bases de données, art. 10 du décret de 1993.

86. C'est-à-dire un livre numérique associant une base de données ou un fichier texte, et un autre document soumis à l'obligation de dépôt légal. Au titre du dépôt légal des documents imprimés, art. 15 du décret de 1993, combiné à l'art. 7 modifié du même décret.

87. Ce support peut aussi être une simple clé USB, un CD, etc. Au titre du dépôt légal des documents imprimés, art. 7 modifié du décret de 1993.

que certaines opérations puissent réellement se développer, ce type de dépôt n'est pas non plus sans difficultés techniques auxquelles il faut porter la plus grande attention.

Chapitre 4 : Les prémices

Dans le cadre d'une réflexion sur le dépôt des *ebooks* à la BnF, on peut penser à deux scénarios schématiques, l'un participant de l'exploitation de l'Extranet éditeurs, et l'autre se rattachant au dépôt légal d'Internet et de la collecte du web. Ces possibilités envisagent la question sous deux angles qui peuvent éventuellement se rejoindre. Plusieurs éléments participent du traitement des *ebooks* dans sa globalité et ne dépendent pas d'une des deux hypothèses en particulier. Nous développerons dans ce chapitre ce que l'on peut considérer comme des prémices à toute réflexion pratique sur le dépôt des *ebooks*, et qui doit donc précéder le choix d'une solution.

Pour comprendre le raisonnement que nous suivons, il faut replacer la problématique du livre numérique dans le contexte concret de la chaîne de traitement du dépôt légal. En ce qui concerne les ouvrages imprimés, cette chaîne se déroule autour du traitement bibliographique du livre après son entrée au département du dépôt légal : il est appareillé, on lui attribue une cote et un code-barres, puis il passe par les services du catalogage, organisés selon la division de la classification Dewey. Après cette étape, il est présent dans le catalogue général de la BnF, il dispose d'une description bibliographique selon des normes précises de catalogage, ce qui lui permet à la fois de trouver une place dans le processus de conservation et de stockage de la bibliothèque (magasins), et d'être facilement retrouvé par un lecteur, par l'intermédiaire des catalogues où il est signalé. La bibliothèque veille à son intégrité et à son bon état physique, indispensables à sa consultation, ainsi qu'à sa valorisation. Elle peut éventuellement numériser l'ouvrage à ces mêmes fins de conservation et de consultation. Dans ce cas, le livre reste présent dans les magasins, mais sa consultation se fait par l'intermédiaire d'un fichier numérique, sorte de nouvelle manifestation de l'ouvrage.

Il est nécessaire de garder cette chaîne à l'esprit lorsque l'on considère la question du traitement du livre numérique. En effet, certaines étapes sont similaires ou parallèles, ainsi que certaines nécessités. Cependant, il va de soi que les termes se posent de manière différente.

I. QUELS FORMATS ?

A. Les formats des *ebooks*

Nous avons déjà abordé la question des formats dans le chapitre 1, nous n'y reviendrons donc que partiellement.

La BnF sera – voire est déjà – confrontée dans sa gestion des livres numériques à différents formats, quelle que soit sa politique en la matière. La multiplicité et la rapide évolution de ces derniers doivent être prises en compte dans la mesure où elles dicteront toute la chaîne de traitement, de la conservation à la consultation. Emmanuelle Bermès attire notre attention sur la nécessité de se pencher sur cette question⁸⁸, lorsqu'elle évoque l'expérience de la bibliothèque universitaire de Cornell et du projet FFMMS (2004-2006)⁸⁹, qui a montré la difficulté, voire parfois l'impossibilité, de récupérer des données dans des formats ou sur des supports obsolètes. Il convient donc de faire le

88. E. BERMES. La conservation des documents numériques. *Les collections électroniques, une nouvelle politique documentaire*, op. cit., p. 209-225.

89. *File format & Media Migration Service*, projet terminé depuis 2006, mais dont le site web est toujours accessible à l'adresse : <<http://www.library.cornell.edu/iris/migration/>> (dernière mise à jour le 2 décembre 2008, consulté le 16 décembre 2010).

nécessaire avant d'avoir atteint ce seuil d'obsolescence, et pour cela, de bien connaître et cerner les paramètres techniques des données à conserver. Dans notre cas, seront à prendre en compte avant tout l'état des supports pour la minorité de livres numériques effectivement disponibles sur support, et surtout les possibilités de migration et de conversion des formats des *ebooks* collectés sous une forme nativement numérique. Nous verrons que le spectre de ces formats peut être très étendu ou beaucoup plus restreint selon la méthode de dépôt choisie, mais qu'elle impose dans tous les cas une diversité nécessitant une vigilance particulière.

B. Les métadonnées

L'objet du dépôt légal ne peut évidemment se résumer à collecter les ouvrages et à les stocker. L'objectif de consultation, à côté de celui de conservation, est à garder à l'esprit de manière prioritaire. Il s'agit de faciliter le classement et le référencement des ouvrages dans les collections, mais également de permettre au lecteur de retrouver exactement l'information qu'il cherche. L'importance de pouvoir clairement et facilement identifier les données et leur nature est soulignée par E. Bermès⁹⁰, en ce qu'elle permet de savoir exactement comment les traiter ; il faut donc prendre en compte ce paramètre au plus tôt dans la gestion des documents.

Le livre numérique, s'il est un objet documentaire singulier, doit trouver sa place dans les collections de la bibliothèque. Les modalités d'accueil des *ebooks* feront l'objet des chapitres suivants, mais quel que soit le mode de dépôt, la conservation et la consultation, voire la valorisation des livres numériques seraient facilitées par l'existence d'une description bibliographique. On peut y réfléchir soit par analogie avec le livre imprimé, soit différemment, selon les obligations qu'imposera(en)t la ou les méthode(s) spécifique(s) de dépôt choisie(s). Cependant, l'objet numérique a cela d'avantageux qu'une part d'automatisation peut être envisagée pour un certain nombre d'opérations.

La description bibliographique des *ebooks* apporte à la bibliothèque les informations nécessaires, qu'elles soient propres à l'« objet livre » (titres, auteur(s), éditeur, nombre de pages, etc.), ou bien qu'elles se rattachent à l'objet numérique en particulier (format, poids, date de création du fichier, etc.). Dans le cas du livre imprimé, c'est la BnF qui crée la notice bibliographique de l'ouvrage. Pour le livre numérique, ces informations, les métadonnées⁹¹, qu'elles soient descriptives (bibliographiques) ou techniques (caractéristiques du fichier), sont liées au fichier et pour la plupart renseignées par l'éditeur ou le diffuseur dès sa création ou en tout cas en amont d'un possible dépôt. Les acteurs de l'édition numérique utilisent en effet pour leurs propres besoins une grande partie des renseignements dont le dépôt légal a besoin, et intègrent ces métadonnées aux fichiers constitutifs des *ebooks* (voir le tableau ci-contre). Dans l'optique de faciliter au maximum le dépôt et d'en réduire les coûts humains, il serait judicieux de récupérer les informations déjà produites par les éditeurs ou les e-distributeurs, que la bibliothèque pourra donc traiter automatiquement, sans rajouter trop d'opérations coûteuses tant pour la BnF que pour les déposants.

Cependant, comme il faudra réfléchir très en amont à la question du format même du fichier déposé, il en est de même pour celui de ces métadonnées. Il existe en effet plusieurs formats pour décrire ces informations. Schématiquement, il s'agirait

90. E. BERMES, La conservation des documents numériques, *op. cit.*

91. Sur les métadonnées du livre numérique, on pourra consulter notamment l'annexe 5 du rapport Patino (*op. cit.*).

aujourd'hui pour le dépôt légal de se positionner entre le format Dublin Core, plus ou moins qualifié, et le format ONIX, qui peut lui aussi être plus ou moins riche⁹².

- Le premier, le format Dublin Core, créé en 1999⁹³, comprend officiellement 15 éléments de description formels, intellectuels et relatifs à la propriété intellectuelle. Il peut être plus ou moins « qualifié », c'est-à-dire que les éléments de base peuvent être raffinés pour gagner en précision. Son avantage réside principalement dans sa simplicité, et il est de fait utilisé très largement par les bibliothèques. La BnF notamment utilise déjà le Dublin Core non qualifié, notamment dans le cadre de Gallica et de SPAR (voir plus bas). Cependant, ce format, dans sa version non qualifiée, est insuffisant pour le dépôt légal, par manque d'informations telles que l'ISBN, la collection, le tirage, etc.
- Le second, le format ONIX⁹⁴, est un fichier XML qui permet de renseigner davantage de données que le Dublin Core (environ 250 éléments). Bien qu'il soit encore un peu balbutiant en Europe, il est très utilisé par les libraires et les e-distributeurs⁹⁵. Gallica traite également des données dans ce format. ONIX présente l'intérêt, dans sa version riche, de réduire très significativement la portion de formulaire nécessaire à remplir pour les éditeurs, puisqu'il précise déjà en grande part les informations nécessaires au dépôt légal.

Quelle que soit la manière de récupérer les fichiers des ouvrages numériques, le traitement des métadonnées qui leur sont attachées doit permettre d'automatiser une grande partie du travail du catalogage et de référencement.

Par ailleurs, cette optique de récupération des métadonnées joue en faveur d'une ouverture du dépôt aux e-distributeurs. En effet, parce qu'ils se trouvent en bout de chaîne, ce sont probablement eux qui sont susceptibles de mettre à disposition de la bibliothèque les fichiers de métadonnées les plus riches et les mieux normalisés. En réduisant le nombre des provenances, cette organisation de masse serait sans nul doute une source d'économies : l'automatisation d'une partie des opérations de vérification et de normalisation serait d'autant facilitée.

92. Voir la liste des éléments de ces deux formats en annexe 4.

93. Ce format est soutenu par une organisation, Dublin Core Metadata Initiative (<<http://dublincore.org/>>).

94. *ONline Information eXchange*. Cette norme est maintenue par un comité international, Onix International Steering Committee.

95. Il faut noter qu'au mois de janvier 2010, Électre, Dilicom et le SNE ont souligné la nécessité d'adopter la norme ONIX for Books, notamment pour l'interopérabilité des fichiers de métadonnées permise par ce format.

	Numilog	Mobipocket	Publie.net	ELG⁹⁶	Eyrolles	Dialogues	Fnac.com	Gallica	DL⁹⁷
Titre	x	x	x	x	x	x	x	x	x
Auteur	x	x	x	x	x	x	x	x	x
Éditeur	x	x	(Publie.net)	x	x	x	x	x	x
Poids/Taille	x	x		x	x	x			(x)
Prix	x	x	x	(non)	x	x	x		x
Prix papier (si nécessaire)	x (1 ^{re} éd.)					x	x		(x)
Format	x	(PRC)	x	x	x	x	x	x	x
Couverture	x	x	x		x	x	x	x	(x)
Descriptif	x	x	x	x	x	x	x	x	(x)
Sommaire/Extrait	x				x			x	(x)
Date de publication/d'édition	x	x	x	x	x	x	x	x	x
ISBN		x	x		x		x		x
Nombre de pages			x				x		x
Collection/Catégorie		x	x		x		x	parfois	x
DRM ou non	(oui)	x	(non)	(non)		x	x		(x)
Tirage									x
Illustrations					x		x		x
Possibilités d'impression et de copier/coller	x	(précisé dans la F.A.Q.)	(oui)				x		(x)
Possibilité de synthèse vocale	x								(x)
Supports de lecture	x								(?)
Feuilletage, avec table des matières	x				x (sans TM)			(x)	

96. Ebookslibresetgratuits.com

97. Les éléments entre parenthèses correspondent à ceux qu'il serait sans doute souhaitable de demander lors du dépôt de livres numériques.

C. La gestion des droits

La gestion des droits sur les *ebooks* représente un des principaux défis techniques et légaux. La plupart des livres numériques actuellement publiés sont munis des verrous que sont les DRM, dont nous avons déjà évoqué l'existence (chapitre 1). Pour le moment, il semble que les éditeurs ne remettent pas vraiment en cause l'application de ces DRM sur les *ebooks*, tant la peur du piratage reste présente. Cependant, des exemples récents pourraient faire évoluer les esprits, tels l'affaire FictionWise, les mouvements d'Adobe sur ce terrain⁹⁸ ou ceux de quelques librairies⁹⁹. Dans le domaine de la musique en ligne, qui devance souvent celui de l'édition numérique, on peut également évoquer le passage éphémère d'une proposition de quasi suppression des DRM dans un des articles de la loi HADOPI¹⁰⁰. Si l'on peut espérer des inflexions sur ce point dans les années qui viennent, on ne doit pourtant pas négliger la question de la gestion des DRM telle qu'elle se présente aujourd'hui. Une partie des *ebooks* que collecterait la BnF dans le cadre du dépôt légal en serait sans doute munie, et pour assurer la pérennité de lecture de ces fichiers, il convient de connaître les obligations qu'imposent les différents formats de DRM.

Dans le cas de fichiers protégés par des DRM, le système le plus sûr actuellement suit un fonctionnement où les *ebooks* sont hébergés de manière sécurisée sur un serveur, que ce soit celui de l'éditeur ou du distributeur. Le client a accès au fichier par une identification à travers le logiciel de lecture, certifiée par la société de gestion des droits. Le fichier téléchargé est ainsi lié à l'appareil de lecture du client, qui ne peut effectuer cette opération qu'un nombre de fois limité pour chacun de ses téléchargements, et donc installer le fichier sur un nombre restreint d'appareils. Ainsi, les DRM n'empêchent pas la lecture du fichier, mais plutôt sa dispersion incontrôlée. Ce système de gestion des droits peut également permettre de gérer d'autres paramètres d'utilisation du fichier (pourcentage autorisé de copier/coller et d'impression par exemple).

Le rapport Patino souligne¹⁰¹ (p. 78) que formats de DRM et formats d'*ebooks* ne sont pas toujours liés dans l'absolu. Il présente les trois formats de DRM identifiés comme standards potentiels :

- l'Ebook Base, qui protège les *ebooks* PRC (qui ne peuvent être protégés que par lui), et qui est hébergé exclusivement sur le serveur de Mobipocket, ce qui met les éditeurs et distributeurs en situation de dépendance avec la filiale d'Amazon (acteur non neutre) ;
- ADEPT et Adobe Content Server, solutions proposées par la société Adobe pour les formats PDF et EPUB, qui garantissent la compatibilité du triplet format/logiciel de lecture/DRM puisque tout dépend de la même société.

Les différences de paramétrage sont considérables d'un système à l'autre : par exemple, l'impression est toujours impossible pour l'Ebook Base, alors qu'elle est entièrement paramétrable avec les solutions d'Adobe.

98. Voir la note sur le blog d'Adobe à l'adresse <<http://blogs.adobe.com/digitalpublishing/2010/11/new-updates-for-ebook-solution.html>> (consulté le 17 décembre 2010).

99. Comme la librairie Dialogues, à Brest, qui a décidé début novembre 2010 de retirer de la vente tous les livres numériques avec DRM disponibles sur son site.

100. Cet article 9 *quater* a été cependant très restreint lors des débats à l'Assemblée nationale.

101. La question des DRM et le détail des plusieurs d'entre eux sont développés dans l'annexe 4 du rapport Patino (*op. cit.*).

La solution que gère le système d'archivage numérique de la BnF est ODRL¹⁰². Ce standard assez souple a été adopté aussi bien par le secteur commercial que par les bibliothèques¹⁰³.

La question des DRM est un axe problématique fort dans la gestion de collections numériques, puisqu'elle conditionne en partie la conservation et la consultation des données : non seulement les DRM peuvent empêcher la pérennité de la consultation par un blocage de l'accès au fichier sur le long terme, mais elles handicapent également la consultation dans un cadre bibliothéconomique (installation sur une multitude de postes). La BnF et son DAV ont déjà rencontré ce problème en ce qui concerne certains documents multimédias, comme les DVD. Comme la loi l'y autorise, elle a pu contourner ces barrières qui empêchent l'installation sur plusieurs postes, mais cette solution pourrait être techniquement plus délicate en ce qui concerne les *ebooks*.

II. LE CIRCUIT DES DONNEES NUMERIQUES A LA BNF

Les *ebooks*, quel que soit leur mode d'entrée dans les collections de la BnF, rejoindront le circuit de conservation des données numériques. Au printemps 2009, la bibliothèque a pleinement lancé le développement de SPAR (système de préservation et d'archivage réparti)¹⁰⁴. Il n'est pas ici question de détailler le fonctionnement de ce système, mais de donner une idée du cheminement qu'adopteront les livres numériques qui entreront à la bibliothèque, et des conditions de conservation de ces documents au sein des collections.

Le marché originel pour la réalisation logicielle du système, lancé en 2007, impliquait notamment les filières de numérisation et de tiers archivage par exemple, mais aussi celles du dépôt légal de l'Internet et du dépôt légal négocié. A été mis en production à l'été 2010 le module de versement pour la filière de numérisation de conservation (Gallica). Les filières de l'audiovisuel, du tiers archivage, ainsi que celle du dépôt légal du Web sont en cours d'instruction, avec une perspective de versement dans SPAR pour la fin de l'année 2011. Les aspects de préservation (migration, émulation notamment) restent également à développer, comme les autres filières (dépôt légal négocié, numérisation de reproduction, archivage BnF, données numériques autres).

SPAR est un vrai magasin numérique intelligent, qui suit le modèle conceptuel de référence OAIS¹⁰⁵. Son principe de préservation est très poussé, c'est-à-dire que non seulement il entrepose les données de manière sécurisée, mais il se charge aussi de surveiller leur présence et leur bon état, d'en effectuer des copies, et surtout de garantir la continuité d'accès à ces fichiers par la réalisation des conversions nécessaires, tout en gardant le fichier original (voir schéma ci-dessous). Ainsi, les fichiers formant un livre numérique seraient mis à jour régulièrement à partir de leur entrée dans le système, en fonction des évolutions liées à leur format, pour maintenir la possibilité de consultation.

102. *Open Digital Rights Management*, est un standard ouvert basé sur le langage XML. Site officiel : <<http://odrl.net/>>.

103. Voir l'étude sur la protection des droits en bibliothèque numérique par Lionel MAUREL. *Panorama des systèmes de métadonnées juridiques et de leurs applications en bibliothèque numérique*. 2006. Disponible à l'adresse : <http://archivesic.ccsd.cnrs.fr/sic_00130222/fr/> (consulté le 20 décembre 2010).

104. Voir la page de présentation du projet SPAR sur le site de la BnF, à l'adresse <http://www.bnf.fr/fr/professionnels/conserver_spar/s.conserver_SPAR_presentation.html> (consulté le 16 décembre 2010), ainsi que l'intervention d'E. Bermès, I. Dussert Carbon, T. Ledoux et C. Ludovici lors du congrès de l'IFLA en 2008, « La préservation numérique à la bibliothèque nationale de France : présentation technique et organisationnelle », disponible à l'adresse : <http://archive.ifla.org/IV/ifla74/papers/084-Bermes_Carbone_Ledoux_Lupovici-trans-fr.pdf> (consulté le 10 décembre 2010).

105. *Open Archival Information System*. On trouvera une explication claire et synthétique de ce modèle sur le blog d'E. Bermès, dans une note datée du 29 novembre 2005, disponible à l'adresse : <<http://www.figoblog.org/document1089.php>> (consultée le 16 décembre 2010).

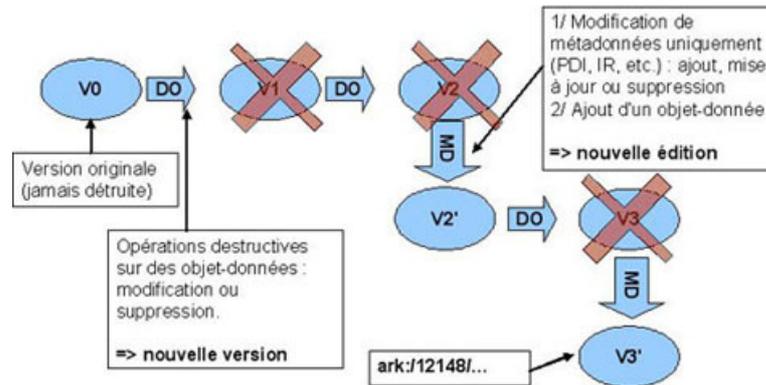


Schéma du cycle de vie d'un paquet d'archive dans SPAR © BnF

SPAR utilise actuellement les formats Dublin Core pour les métadonnées de description et ODRL pour les licences d'utilisation, sachant que le projet tient à s'engager dans des standards faisant autorité. La gestion des formats d'entrée répond au domaine de la politique de versement¹⁰⁶, et participe de quatre catégories intégrées par le système et décrivant la responsabilité du producteur pour la préservation future de ses données :

- stocké : format dont les caractéristiques techniques sont inconnues (non identifiées), et pour lequel la seule préservation du flux de bits est assurée ;
- identifié : format dont les caractéristiques techniques sont connues (identifiées dans un répertoire de formats), mais pour lequel ni la migration ni l'émulation ne sont envisagées ; un format identifié devient maîtrisé ou connu si un tel plan est mis en œuvre ;
- connu : format identifié pour lequel la BnF détient au moins un outil de référence, connaît ses utilisations, suit les évolutions, et pour lequel la BnF a défini un plan soit pour le transformer en un format maîtrisé, soit pour l'émuler ;
- maîtrisé : format connu pour lequel la BnF détient la documentation publiée et au moins un outil de référence, dont elle suit les évolutions et pour lequel elle a défini des contraintes spécifiques avec les producteurs.

La catégorie des formats d'entrée influe sur la politique de préservation puisqu'un format doit être maîtrisé pour pouvoir bénéficier d'une stratégie de migration. On voit là l'importance de la réflexion à avoir en aval sur les formats, et sur les DRM, qui conditionnent l'accès au fichier et aux données.

À côté de cet entrepôt se trouvent les serveurs de consultation, aujourd'hui principalement destinés à Gallica, qui permettent également de consulter les *ebooks* déjà présents dans la bibliothèque numérique de la BnF. La recherche se fait alors par le moteur Lucene¹⁰⁷, à la fois sur les champs des notices bibliographiques en Dublin Core¹⁰⁸ et dans le plein-texte lorsqu'il est disponible. L'accent est mis sur la consultation

106. Voir E. BERMÈS et autres, La préservation numérique..., art. cité, p. 7.

107. <<http://lucene.apache.org/>>.

108. Ces notices en Dublin Core non qualifiées sont contenues dans l'entrepôt « OAI-NUM » de la BnF, qui signale les documents provenant des collections numériques. Pour une description plus détaillée de cet entrepôt, voir l'article le concernant dans l'Espace professionnel de publication sur le document numérique à la BnF, disponible à l'adresse <<http://bibnum.bnf.fr/oai/index.html>> (consulté le 16 décembre 2010).

et la mise à disposition des données, et non sur leur conservation, comme c'est le cas dans SPAR.

Quel que soit le scénario de collecte ou de dépôt retenu, les *ebooks* acquis par dépôt légal se trouveront de toute façon conservés dans SPAR, puisqu'il s'agit de données numériques. En ce qui concerne l'interface de consultation, elle dépendra sans doute en grande partie de la manière dont ces documents arriveront dans les collections de la BnF, mais aussi des choix de valorisation et du statut accordé à ces documents – c'est une question que nous retrouverons plus bas, dans l'étude des deux possibilités.

III. LA MISE EN PLACE D'UNE INTERFACE DE CONSULTATION

L'un des défis du traitement des *ebooks* déposés est leur mise en consultation, même considérée hors de la question des DRM et du dépôt à strictement parler. La multiplicité des formats et des outils de lecture implique à la fois une réflexion en aval, mais aussi une organisation technique dédiée. Par exemple, pour lire des livres numériques au format PRC, il faut utiliser le logiciel Mobipocket Reader. Certes, ce logiciel est gratuit et facilement accessible, mais il implique une installation et une mise à jour sur tous les postes où la consultation des *ebooks* sera possible. Or aujourd'hui, seule l'interface pour accéder à Internet (navigateurs de base), ainsi que le logiciel d'Adobe (ADE) sont disponibles depuis les postes du Rez-de-Jardin (pas de suite bureautique par exemple), et cela peut rendre la consultation de certains documents difficile voire impossible. Si l'on veut mettre les *ebooks* collectés via le dépôt légal à disposition des chercheurs, comme la loi le prescrit, il est nécessaire de trouver une solution technique dans ce domaine. Pour cela, on peut envisager plusieurs solutions.

Les archives du Web peuvent offrir une première interface de lecture. Le cas du feuilletage en ligne paraît assez simple à gérer. En effet, si le site où se trouve l'*ebook* a été collecté, il est vraisemblable que les possibilités de lecture en *streaming* resteront les mêmes, si le robot moissonneur a été bien configuré pour aller chercher jusqu'au fichier lu. Dans ce cas, l'ouvrage sera disponible assez simplement dans l'interface de consultation des archives du Web (voir chapitre 6).

En revanche, pour les livres numériques qui doivent être téléchargés, le problème se pose en de tout autres termes. Pour le résoudre, on pourrait penser à la mise en place d'espaces numériques de travail (ENT) pour les chercheurs habilités, à la manière de ce qui se fait dans certaines bibliothèques universitaires. Les lecteurs disposeraient alors des logiciels nécessaires pour lire des *ebooks*, à travers une interface dédiée : ADE, Mobipocket Reader, des émulateurs éventuels pour les formats plus spécifiques liés à des supports. On peut également envisager l'exploitation du profil lecteur qui permet à chaque usager de la bibliothèque muni d'une carte de faire ses réservations, d'avoir un panier de notices sur le catalogue, etc. Ce profil pourrait donner accès à l'interface de lecture des *ebooks* et aux logiciels de lecture spécifiques.

Une troisième hypothèse serait de mettre les *ebooks* conservés par la BnF en accès sur Gallica. Cette solution serait peut-être plus simple à mettre en place, puisque déjà en partie effective, même si elle demanderait sans doute une réflexion technique d'élargissement des possibilités de la bibliothèque numérique de la BnF.

Les *ebooks* pourront donc être mis à disposition du lecteur à travers différentes interfaces, qui dépendront notamment du mode d'arrivée de l'ouvrage dans les collections. Nous verrons que le référencement de ces *ebooks* pourra plus ou moins être développé. Il importe cependant de pouvoir filtrer l'utilisation de ces logiciels et l'accès aux *ebooks* puisque la loi n'autorise l'accès aux documents du dépôt légal que pour les personnes autorisées, en l'occurrence, les lecteurs du Rez-de-Jardin.

En ce qui concerne les supports physiques de lecture, les liseuses, il faut également insister sur la tradition de conservation des supports de lecture des données multimédias de la BnF. Dans la lignée de la collection Charles Cros, l'établissement collecte ainsi les appareils d'enregistrement et de diffusion en lien avec les documents multimédias qu'elle conserve par ailleurs. L'édition électronique est elle aussi concernée par cette collecte depuis 1992¹⁰⁹, gérée par le département de l'Audiovisuel. Dans ce cadre a été entreprise la constitution d'une collection de liseuses, dont les plus anciennes sont pour le moment le premier modèle commercialisé par Sony en 2006 et l'Iliad d'Irex, même si l'arrivée de liseuses de la première génération, comme le Rocket eBook, devrait se faire sous peu¹¹⁰. Il n'existe pas d'appui législatif pour justifier un dépôt légal de ces sortes d'appareils¹¹¹, il s'agit donc pour la bibliothèque de les acquérir. Si cette politique se place encore aujourd'hui dans une optique systématique, les responsables se posent des questions face à l'inflation des modèles en circulation. Il est moins ici question des livres numériques en eux-mêmes, puisqu'on pensera de toute façon à les recopier sur des supports de masse pour des raisons de conservation¹¹², mais de sauvegarder une partie de l'environnement de ces *ebooks*, selon la logique de la collection Cros.

À côté de cette optique de conservation du support, on pourra aussi évoquer la possibilité d'utiliser ces liseuses dans le cadre de la consultation des *ebooks*, notamment ceux qui pourraient être dépendants de plateformes de diffusion hors Web. Ces appareils de lecture pourront éventuellement aussi être communiqués, selon des conditions qu'il conviendra de déterminer (demande motivée du lecteur), tout en gardant bien à l'esprit ce qu'un tel système aurait de difficile à mettre en place, à la fois à cause de la diversité croissante de ces appareils et des modalités pratiques de prêt (prêt de l'exemplaire de conservation ? deux exemplaires ?).

109. M. COHEN-HADRIA. Étonnante collection Charles Cros. Paris : BnF, octobre-décembre 2006. *Chroniques de la Bibliothèque nationale de France*, n° 37. Disponible à l'adresse : <http://chroniques.bnf.fr/archives/decembre2006/numero_courant/coulisse/charles_cros.htm> (consulté le 16 décembre 2010).

110. Selon un échange avec Élodie Bertrand, responsable du dépôt légal des documents multimédias à la BnF.

111. Mis à part le cas où des liseuses seraient fournies avec du contenu intégré (Code du patrimoine, art. L131-2), mais d'après E. Bertrand, ce cas n'a pas encore mis en pratique.

112. E. BERMES. La conservation des documents numériques, *op. cit.*, p. 211.

Chapitre 5 : Un dépôt par l'Extranet éditeurs

La première proposition que l'on peut élaborer pour gérer le dépôt légal des *ebooks* utilise un outil déjà existant, l'Extranet éditeurs. Il s'agit là de se placer dans une démarche se rapprochant de celle utilisée pour les dépôts analogiques, ce qui induit une participation active de l'éditeur, ou plus largement du déposant, dans le processus.

I. L'EXTRANET EDITEURS

Comme nous l'avons déjà rapidement mentionné, l'Extranet éditeurs est un service mis à disposition des déposants éditeurs par le dépôt légal dans une optique de simplification par la dématérialisation des formalités. Le déposant doit se connecter sur une plateforme sécurisée, après avoir obtenu un identifiant et un mot de passe. Ceux-ci sont délivrés par la BnF après une inscription qui consiste en la fourniture des coordonnées de l'éditeur, personne physique ou morale, ainsi que du nom du responsable du compte. Ceci permet à la bibliothèque de s'assurer que le demandeur est bien un producteur de documents¹¹³ et qu'il n'est pas déjà enregistré dans la base. Le 2 novembre 2010, 3.091 éditeurs s'étaient inscrits sur l'Extranet éditeurs¹¹⁴. Cette procédure d'inscription doit donc être validée par la BnF, et les conditions d'utilisation acceptées par le déposant. Ces conditions¹¹⁵ sont régulièrement mises à jour, selon les évolutions de l'Extranet. Elles décrivent notamment le service, ses conditions d'accès, ses enjeux techniques et intellectuels (sécurité, publication des données, etc.).

La page d'accueil du service annonce clairement les trois services mis en avant actuellement :

- la déclaration en ligne des documents parus ou à paraître,
- le suivi de leur traitement en temps réel,
- l'accès aux notices bibliographiques de ces documents.

De plus, il permet l'actualisation des informations (coordonnées, contacts) de part et d'autre, favorisant ainsi un lien direct toujours nécessaire entre les déposants et les services du dépôt légal. La correspondance et les échanges d'informations ont lieu par voie dématérialisée. L'accès se fait par protocole sécurisé (https) car certaines informations fournies par le déposant doivent rester confidentielles, comme le chiffre de tirage du document ou bien le nom réel d'un auteur écrivant sous un pseudonyme.

Pour tout dépôt, il est donc nécessaire pour le producteur (au sens large) de remplir un formulaire¹¹⁶, étape simplifiée et rendue plus rapide grâce à l'Extranet¹¹⁷. Dans ce formulaire, il doit indiquer :

113. Si le demandeur n'est pas déjà inscrit dans la base de la BnF, c'est-à-dire qu'il dépose déjà de manière traditionnelle, ou dans celle de l'AFNIL, des précisions lui sont demandées par mail. Les demandes d'inscription sont traitées dans la journée.

114. 1.393 avaient effectivement déposé au moins un imprimé grâce à l'Extranet, selon un entretien avec Anne Bommier-Chasles, responsable de l'Extranet.

115. Page des conditions d'utilisation de l'Extranet : <https://depotlegal.bnf.fr/cond_utilisations.html>. La dernière mise à jour valable pour cette étude est celle du 15 novembre 2010.

116. Voir des exemples de ces formulaires en annexe 3, pour les dépôts d'imprimés, de documents multimédias et de documents informatiques. Ces formulaires de dépôt sont également accessibles sur le site de la BnF, dans la rubrique du dépôt légal.

- son identité : nom (ou raison sociale), adresse, téléphone, courriel, fonction dans la production du document, nom du correspondant pour le dépôt légal, date, lieu et signature ;
- et les données concernant le dépôt en lui-même : titre, nom et prénom de l'auteur, pseudonyme éventuel, informations sur la collection et le responsable de publication, ISBN/ISSN, format (taille, nombre de pages, façonnier, reliure), tirage (chiffre du premier tirage), statut de l'édition (nouveau, nouvelle édition), prix de vente et date de mise à disposition du public, éventuellement informations techniques (nature du support, système de protection, documentation, environnement requis).

Le fait de renseigner ce formulaire en ligne ne dispense pas le déposant de devoir l'imprimer et le joindre à son dépôt physique, généralement effectué par envoi postal¹¹⁸. Mais il permet d'accélérer la procédure de traitement, notamment par des possibilités de pré-remplissage de certains champs du formulaire (collection, livre ou périodique, imprimeur). Il donne également accès au suivi de l'évolution du statut du document, qui se décompose en quatre étapes : déclaration en cours, validée, dépôt enregistré, document catalogué ; chacune de ces étapes offre à l'éditeur des liens vers des documents spécifiques (déclaration, notice du catalogue général). Cela permet une relation plus transparente avec les déposants, qui peuvent avoir rapidement connaissance de l'avancée du traitement de leurs dépôts.

Les données bibliographiques que fournit le déposant par le biais du formulaire sont utilisées pour la publication de la *Bibliographie nationale de la France* et, après enrichissement par les services du catalogage, versées dans le catalogue général de la BnF. A été mis en place très récemment le service « Nouveautés éditeurs »¹¹⁹, qui devrait trouver sa pleine activité à partir de janvier 2011. La création de ce service a nécessité l'ajout dans le formulaire d'une demande d'autorisation de publication de certaines des informations fournies : le déposant précise s'il donne son accord à la publication automatique d'une partie du contenu du formulaire¹²⁰, dès que la déclaration est enregistrée, avant même la validation par la BnF. Si oui, paraît sur le site dédié une notice avec les informations en question et une date de parution, et éventuellement une image de la couverture de l'ouvrage¹²¹, assortie d'un descriptif du contenu. Les données sont actuellement visibles pendant trois mois après la date de publication de l'ouvrage. Ce service présente un intérêt pour le grand public, mais surtout pour les professionnels (libraires, acquéreurs, éditeurs), afin de leur offrir une vue des nouveautés de l'édition en temps réel, avec les futures parutions. Un service tel que Nouveautés éditeurs est également une motivation pour les déposants à affiner la qualité des renseignements fournis au départ : ce qu'ils inscrivent dans le formulaire est exactement ce qui sera annoncé sur le site. Par ailleurs, la publicité qu'apporte ce service permet d'inciter certains déposants à utiliser l'Extranet, voire plus largement à effectuer leur dépôt légal ; c'est notamment le cas dans le domaine de l'auto-édition, qui trouvera sans doute là un terrain de visibilité idéal.

117. Selon l'enquête de satisfaction des éditeurs sur l'Extranet (2010), c'est même ce critère de la rapidité offerte par le service qui emporte tous les suffrages. Une synthèse des résultats de cette étude est disponible à l'adresse : <http://www.bnf.fr/fr/professionnels/anx_pro_actu/a.etude_extranet_editeurs.html> (consulté le 20 décembre 2010).

118. Selon l'étude de satisfaction déjà mentionnée, les éditeurs n'attendent visiblement plus que d'être libérés de cette obligation d'envoyer un exemplaire papier du formulaire.

119. <<http://editions.bnf.fr/nouveautes/sommaire.htm>> (consulté le 19 décembre 2010).

120. Les données publiées et celles qui ne le sont pas se trouvent précisées dans les conditions d'utilisation de l'Extranet.

121. Cet élément est encore expérimental.

II. QUELLES POSSIBILITES POUR LE DEPOT LEGAL DES EBOOKS ?

A. Les déposants

Comme nous l'avons déjà mentionné à propos des formats de fichiers, le dépôt légal des *ebooks* implique une réflexion sur l'ouverture du statut de déposant aux e-distributeurs. Cette solution offre des avantages conséquents en termes de traitement de masse et de normalisation des dépôts, sans compter la réduction du nombre des interlocuteurs pour la bibliothèque. Il faut cependant obtenir pour cela l'accord des éditeurs qui sont ceux concernés au premier chef par l'obligation de dépôt légal. La négociation de cet accord pourrait se faire au niveau de la BnF, ou bien à celui des e-distributeurs, qui obtiendraient un mandat pour effectuer les dépôts de la part des éditeurs avec lesquels chacun travaille. Une entente directe entre e-distributeurs et éditeurs selon le principe d'une délégation serait bien sûr plus simple pour la bibliothèque. Ceci permettrait un dialogue renforcé et bienvenu entre les acteurs de la chaîne du livre numérique, dans lequel les organisations professionnelles (type SNE) pourraient avoir un rôle important à jouer. Même si la BnF doit donner une définition claire du cadre de l'accord, celle-ci n'aurait alors pas à gérer l'aspect négociation ; un certain suivi sera cependant nécessaire pour éviter des pertes (en cas de faillite, de cessation d'accord entre deux acteurs, etc.). Il faut souligner qu'une telle organisation ne serait pas complètement inédite : pour les périodiques papier, il est courant que ce soit le distributeur ou l'imprimeur qui dépose par délégation, d'où un encart « Déclaration pour le compte de » dans le formulaire de dépôt. Il faut noter que même dans ce cas de délégation, l'éditeur reste en droit celui qui effectue le dépôt, aucun changement législatif n'est donc à prévoir.

En ce qui concerne d'abord la procédure d'inscription à l'Extranet, il faudra demander au déposant des informations supplémentaires, comme l'adresse du site web qu'il gère par exemple, afin de suivre la logique de vérification du statut de producteur d'*ebooks*¹²². Dans le formulaire de dépôt, ces renseignements supplémentaires trouveraient également leur place dans la rubrique consacrée au déposant ; en s'inspirant du formulaire de dépôt des documents informatiques¹²³, on demanderait la précision de la fonction du déposant.

Le service des Nouveautés éditeurs peut être exploité dans le cadre des *ebooks*. L'avantage serait pour l'éditeur ou le e-distributeur de disposer d'une base des nouveautés éditoriales et des à paraître avec un lien direct vers le service de vente ou le site de l'éditeur. L'exploitation de ce service et les avantages qu'il peut amener à certains éditeurs, notamment en termes de publicité, pourront jouer en faveur d'un dépôt difficile à circonscrire et à appréhender qui est celui de l'auto-édition. Il en est de même pour la présence dans Gallica d'*ebooks* encore sous droits et payants, à l'accès restreint, mais bénéficiant d'une notice enrichie (table des matières, couverture) et de la recherche plein texte. L'expérimentation actuelle pourra favoriser une démarche où la BnF fournit aux éditeurs un levier de visibilité et de valorisation en échange de la fourniture de données fiables sur les ouvrages. Cet échange de bons procédés peut constituer un encouragement puissant à une coopération rafferme.

122. Cette adresse de site web peut correspondre à plusieurs statuts : le site institutionnel de l'éditeur et/ou du distributeur, celui qui héberge le catalogue du déposant, ou celui qui héberge les fichiers en question. Il s'agit de décider à quel niveau de granularité on fixe la demande de précision de ce site. Cette précision revêt encore davantage d'importance dans un scénario de collecte par le dépôt du web (chapitre 6).

123. Dans la rubrique « Identification du déposant » sont proposées comme fonctions : éditeur, producteur, commanditaire et distributeur — cette dernière pouvant inclure celle de e-distributeur, dans notre cas.

B. Le dépôt

Le dépôt des *ebooks* par le biais de l'Extranet se ferait à partir du formulaire de dépôt déjà existant. Celui-ci doit s'adapter au traitement de livres numériques, bien qu'il soit relativement facile de le calquer sur celui des imprimés, en rajoutant des éléments inspirés des autres déclarations (périodiques, documents informatiques, documents multimédias).

Il conviendrait notamment de rajouter la notion de format numérique, selon le choix de la BnF en la matière. Dans ce scénario, la bibliothèque devrait idéalement choisir quelques-uns des formats qui existent aujourd'hui pour les *ebooks* et de ne proposer le dépôt que dans l'un des formats maîtrisés par la BnF, même si cela réduit la marge de manœuvre des déposants. Ceux-ci pourraient ainsi remettre leur document en formats PDF ou EPUB, voire PRC, qui sont les plus répandus sur les plateformes de vente. Ce choix restreint a l'avantage de simplifier le traitement ultérieur des fichiers en réduisant le nombre de formats à gérer et en se limitant aux formats les mieux maîtrisés du point de vue de la conservation – même s'il restera toujours des incertitudes en la matière. Il faut garder à l'esprit l'automatisation d'une certaine partie du circuit en aval, qui est facilitée par un panel réduit de formats à traiter et à suivre. Un contrôle automatique vérifierait le format du document déposé et s'il correspond bien à celui déclaré. Circonscrire les formats de dépôt permettra une meilleure prise en charge au sein de SPAR : selon la typologie des formats que nous avons vu précédemment (chapitre 4), il serait souhaitable à terme de n'accepter que des formats au moins connus, ce qui permettrait d'en suivre les évolutions. Actuellement, l'EPUB semble peu à peu s'imposer comme un format propre à devenir standard ; il faudrait également considérer le PDF, extrêmement répandu.

Un certain nombre d'autres éléments doivent être ajoutés au formulaire, comme la présence et le type éventuel de DRM qui accompagnent le fichier¹²⁴, l'existence d'une édition papier et le statut de l'*ebook* par rapport à celle-ci (contenu identique ou enrichi, publication antérieure, simultanée ou postérieure, etc.). Le traitement des métadonnées pourrait également se faire à ce moment, puisqu'on a vu qu'avec certains formats (ONIX riche notamment), le déposant n'aurait quasiment pas besoin de remplir manuellement le formulaire de dépôt.

Le dépôt des *ebooks* par l'Extranet nécessiterait la création de liens entre les différentes bases de la BnF, comme celle de Gallica. Là encore, de tels liens existent déjà, notamment la possibilité pour le déposant d'accéder directement à la notice de son document traité grâce à la présence de l'identifiant ARK¹²⁵ de la notice du catalogue général. Peut-être pourrait-on penser à des liens similaires vers les fichiers d'*ebooks* conservés dans SPAR, comme il est actuellement envisagé d'intégrer le dépôt de l'image de la couverture dans le formulaire. Le déposant pourrait donc trouver sur son interface personnelle non seulement l'état d'avancement du traitement de son dépôt et la notice de celui-ci, mais également un lien direct vers l'ouvrage conservé.

Une fois le dépôt réalisé et contrôlé automatiquement, les métadonnées qui lui sont rattachées pourront être également traitées de manière semi-automatique et exploitées pour un catalogage plus rapide, même si une vérification et un enrichissement d'au moins une partie de ces données seront sans doute nécessaires. Si le document ne diffère

124. Même s'il faut insister sur le fait que la bibliothèque aurait tout à gagner à des précisions pédagogiques soulignant la difficulté de traiter des documents équipés de ces DRM.

125. Le système ARK (ARchival Resource Key) a été choisi par la BnF pour créer des identifiants pérennes, qui sont des adresses Internet particulières qui permettent de citer un document numérique à l'aide d'un lien hypertexte, tout en ayant la garantie que ce lien ne risque pas de changer. Chaque notice dans le catalogue général a ainsi un numéro ARK. On trouvera plus d'informations sur les identifiants pérennes sur l'article de bibnum.bnf.fr qui leur est consacré, disponible à l'adresse <<http://bibnum.bnf.fr/identifiants/index.html>> (consulté le 19 novembre 2010).

pas vraiment d'un document analogique déjà présent dans les collections, on pourra se contenter d'une exemplarisation ou d'une modification à la marge de la notice du document préexistant dans le catalogue¹²⁶ ; sinon, s'il n'existe qu'en ligne ou que les versions numériques et papier diffèrent sensiblement, une nouvelle notice doit être créée pour l'*ebook*. Aujourd'hui, l'orientation générale reste difficile à prendre, en raison de la complexité à appréhender les volumes en jeu (peu de chiffres disponibles). Il est évident que le taux d'inédits numériques pèsera sur les efforts à fournir en termes de catalogage. Le marché du livre numérique étant encore mal connu et en pleine évolution, une certaine flexibilité est nécessaire, alors même que cette question du catalogage et de la gestion des flux est centrale dans notre sujet.

Le fichier de l'ouvrage lui-même sera versé dans SPAR, et dans Gallica pour laquelle il fera l'objet d'une indexation plein texte¹²⁷. Il s'agirait de deux fichiers distincts (copies), car l'exemplaire de consultation de Gallica ne peut a priori être celui de conservation dans SPAR, avec une nécessité de tenir l'exemplaire de consultation à jour. Ainsi, deux modes de recherche permettraient d'accéder à l'*ebook* et à son contenu, avec une intéressante possibilité de rebond entre les deux : à partir du catalogue, comme les autres types de documents conservés par la BnF qui y disposent d'une notice, ou à partir de Gallica, grâce à une recherche plein texte parmi les documents numérisés. Si l'éditeur donne son accord, le lecteur navigant sur Gallica hors du Rez-de-Jardin pourra feuilleter quelques pages de l'ouvrage sous droits. Il sera redirigé vers un site vendeur s'il ne peut venir directement consulter le document à la bibliothèque, à la manière de ce qui se fait déjà pour les *ebooks* sous droits disponibles aujourd'hui dans Gallica.

C. Nécessités

1. L'information en direction des déposants

À ce point se repose la question de la définition de l'*ebook*. En effet, il serait efficace, voire attendu par les déposants, de proposer une description *a minima* des objets numériques pouvant faire l'objet d'un tel dépôt¹²⁸. Actuellement, les pages du site bnf.fr¹²⁹ renseignant les déposants sur ce sujet reprennent la définition sommaire que nous avons mentionnée plus haut¹³⁰. Il est précisé aujourd'hui que ce dépôt se fait dans le cadre du dépôt légal de l'Internet et que le déposant n'a pas de démarche active à effectuer. Il est évident que si une solution met en œuvre l'Extranet, ces précisions devront être modifiées.

Il est également fait mention que le dépôt d'un livre numérique ne se substitue pas à celui de la version papier. On a vu avec le dépôt de substitution que cette règle pouvait être remise en question. Il faudrait, en ce qui concerne le livre numérique, prendre une position arrêtée.

126. Dans ce cas, il faudra veiller à la date d'arrivée des documents : la notice pourra renvoyer indifféremment au document papier ou au document numérique selon celui qui aura été traité en premier par les services du dépôt légal.

127. Cela si Gallica est retenue comme interface de consultation ; cela nécessiterait sans doute un développement des possibilités de Gallica, au vu des gros volumes de données à accueillir.

128. En effet, dans l'étude sur l'Extranet mentionnée plus haut, les non utilisateurs citent notamment parmi les raisons de leur non-usage du service, le manque d'information (pour 21 %).

129. Les précisions sur les livres numériques se trouvent à la fois sur la page du dépôt « Livres » et sur celle du dépôt des sites Web.

130. « Le dépôt légal concerne les e-books ou livres numériques, termes utilisés pour désigner un objet numérique ressemblant en partie à une monographie imprimée sur papier et diffusé en ligne. » Avec la précision : « Seul le contenu (le texte numérique ainsi que les fonctions d'annotation, les outils interactifs, etc.) est visé par le dépôt légal, et non l'outil de lecture ou tablette. »

Il faudrait également préciser davantage que le dépôt d'*ebooks* dont la lecture dépend d'une liseuse se fait par le département de l'Audiovisuel (dépôt sur support), et que de toute façon, la liseuse en elle-même n'est pas soumise à une obligation de dépôt. La solution pour éviter une possible dissémination des dépôts de livres numériques, et donc potentiellement des erreurs et des pertes de temps, serait de demander aux déposants de tels ouvrages le dépôt par l'Extranet, de manière classique, mais dans un format neutre, même s'il ne correspond pas à celui utilisé par la liseuse. Cela simplifierait à la fois la procédure de dépôt et celle de conservation, même s'il faudrait mettre en place des accords avec les déposants concernés, en raison du côté législatif bancal de cet arrangement¹³¹.

Parallèlement à l'information sur le site Web de la bibliothèque, l'Extranet dispose d'autres moyens pour communiquer avec les éditeurs¹³², comme le prospectus¹³³. Peut-être serait-il souhaitable, au moins pendant la mise en place du service, d'enrichir ce dernier, voir d'en élaborer un nouveau. Des mails pourront également être envoyés aux éditeurs visés pour leur faire part de l'existence de ce service. L'information en direction des éditeurs est capitale, non seulement parce qu'elle permet de les mettre au courant de l'existence et de l'évolution du service, mais également pour répondre à toutes les questions techniques qui ne manqueront pas d'apparaître (sur les formats, les liens avec la collecte du Web). Les difficultés pour atteindre un maximum d'éditeurs demanderont sans doute une certaine diversité des moyens d'information, et donc un coût indéniable pour la bibliothèque, mais celle-ci a tout à gagner à faire parler de ses initiatives dans un secteur encore en développement.

2. Une mise en place phasée

En raison du caractère expérimental de cette approche du dépôt légal, il serait prudent d'adopter une démarche progressive. D'une part, en ce qui concerne le document déposé lui-même, le maintien du dépôt d'une version papier devra se faire en parallèle avec la version numérique, dans le cas des ouvrages qui sont identiques sous les deux versions, selon le principe du dépôt légal qui privilégie le support et non l'œuvre. Actuellement, l'éditeur envoie à la BnF deux exemplaires papier, dont l'un est pilonné après traitement du dépôt, s'il n'est pas utilisé par le service des échanges. La version numérique pourrait alors remplacer le second exemplaire aujourd'hui exigé par le dépôt légal ; il s'agit cependant d'être prudent, puisqu'une telle modification ne pourra se faire sans une révision des textes législatifs¹³⁴. Ensuite, on pourra envisager une extension de l'expérience en un dépôt de substitution tel qu'il existe déjà pour les affiches grand format par exemple¹³⁵. Il s'agit de bien situer progressivement tous les enjeux et de ne pas se laisser déborder par les incontournables difficultés techniques que l'ouverture d'un tel dépôt va engendrer. Ces premières étapes devront s'accompagner de la mise en place d'un processus de récolte des *ebooks* qui n'existent que sous forme numérique et qu'il devient de plus en plus nécessaire de recueillir, pour des raisons patrimoniales.

131. En effet, ces documents seraient à la base soumis à l'obligation du dépôt sur support, mais en pratique, ne seraient pas déposés sous la forme soumise. Peut-être serait-il possible de rapprocher cette opération du dépôt par substitution ?

132. Selon l'enquête de satisfaction, 73 % des éditeurs ont été informés de l'existence et du fonctionnement de l'Extranet par la promotion réalisée par la BnF, ce qui montre une certaine efficacité de celle-ci dont pourra profiter le dépôt des *ebooks*. Cependant, les auto-éditeurs s'informent généralement eux-mêmes sur les services existants (43 %), ce qui rend la mise à jour de la page Web indispensable.

133. Disponible à l'adresse <http://www.bnf.fr/documents/brochure_extranet_depot_legal.pdf>.

134. Décret de 1993 relatif au dépôt légal, modifié en 2006.

135. Voir le chapitre 3 pour le contexte législatif du dépôt de substitution.

De manière encore plus cruciale, cette approche progressive devra également se faire à travers le choix des interlocuteurs, car c'est de leur investissement, dans un dépôt actif comme celui-ci, dont dépend en partie la bonne réussite du projet. Choisir un panel restreint d'éditeurs et/ou de e-distributeurs pour commencer l'expérience est sans doute stratégiquement plus intéressant et plus simple. La BnF travaille d'ores et déjà avec des e-distributeurs et des éditeurs intéressés dans le domaine du livre numérique, à travers Gallica. Il est logique de commencer par s'entendre avec ces acteurs pour réfléchir au dépôt des *ebooks* : ils connaissent déjà l'établissement et celui-ci a entamé un dialogue avec eux sur ce domaine précis. Il est important de garder à l'esprit que certains déposants sont plus ou moins concernés par l'Extranet ; ceux qui le sont davantage ne pourront que fournir aux débuts de l'expérience un terreau favorable. Une base de volontariat semble incontournable dans les premiers temps, afin de bénéficier d'interlocuteurs motivés et à même de comprendre une certaine impression de double dépôt légal¹³⁶ dans une phase d'expérimentation. Les résultats de cette première phase permettront de réviser les conditions du dépôt numérique afin de l'affermir avant de le proposer à un plus grand nombre de déposants.

Il semble bien que le caractère progressif de la mise en œuvre de ce scénario soit central pour sa réussite. Au vu des flous et du contexte mouvant, il faut s'engager avec prudence et flexibilité dans ce projet, et seule une expérimentation avec un nombre limité de producteurs, sur une offre circonscrite d'*ebooks* permettra une mise en place réfléchie et adaptée.

3. Coûts et moyens

Aujourd'hui s'occupent de l'Extranet deux chargées de mission à temps plein. Le service concerne également trois autres personnes à temps partiel, à raison d'une demi-journée par semaine chacune. Certains autres agents du département du dépôt légal peuvent être ponctuellement impliqués dans l'Extranet, mais ce travail se substitue alors à celui qu'ils effectuent pour les dépôts papier¹³⁷. Au DSI, deux informaticiens gèrent les aspects techniques de l'Extranet (à plus de 50 % de leur temps), en coordination étroite avec les deux chargées de mission ; les développements nécessaires sont confiés à une entreprise (ce qui occupe 1,5 ETP).

Afin de mener le projet d'extension de l'Extranet, on peut compter sur un certain nombre d'éléments qui sont déjà pris en compte dans le fonctionnement du service et des applications :

- le développement de fonctionnalités de chargement de fichiers numériques est déjà prévu dans la feuille de route de l'Extranet¹³⁸ ;
- la refonte du catalogue général peut également englober les évolutions induites par la prise en compte des *ebooks* (lien vers le document par exemple).

Du point de vue informatique, il s'agit non seulement de réfléchir sur la mise en place initiale du service étendu aux *ebooks*, avec ses différentes dimensions¹³⁹, mais

136. À la fois papier et numérique, incontournable dans un premier moment.

137. Il s'agit avant tout de membres des services d'entrée et du catalogage.

138. On peut évaluer la charge supplémentaire à 2 ETP extérieurs.

139. Dont le développement des fonctionnalités de l'Extranet (formulaire de dépôt, interface avec les éditeurs), le dégagement d'espaces de stockage avant le versement des ouvrages dans SPAR et sur les serveurs de consultation de Gallica, les outils de vérification automatique et semi-automatique des données fournies. Pour plus de liberté et de marge de manœuvre future, on pourra considérer la création dans SPAR d'une nouvelle filière, à côté de celles du dépôt négocié et du dépôt légal du Web ; le choix du développement de cette nouvelle filière engendrerait des coûts supplémentaires, dont la consultation d'un ou

également à l'entretien et au suivi de ces structures. Sur le long terme, le travail de veille (sur la mise à jour des formats par exemple), de gestion des réclamations (de la part des déposants et des services de la BnF), de contrôle qualité doit bien être gardé à l'esprit.

En supplément des effectifs déjà concernés, on peut estimer les effectifs supplémentaires à un chef de projet (A) parmi l'équipe de l'Extranet, 7 gestionnaires de collections (B) parmi les équipes du DDL et 4 ingénieurs de production au DSI (A), autant de personnes qui ne travailleront pas sur le projet à temps plein (estimation à 5,5 ETP). Selon la solution organisationnelle choisie, ces personnes pourront travailler au sein d'une équipe particulière (détachement de la filière de l'imprimé) ou, de manière sans doute moins satisfaisante, réparties parmi les différents services du dépôt légal (calque sur la filière de l'imprimé). L'automatisation doit être privilégiée à toutes les étapes, dans un souci d'économie des ressources humaines. De plus, au fur et à mesure que la procédure se mettra définitivement en place, on peut escompter sur un certain taux de remplacement du traitement papier par le traitement numérique, sur le long terme cependant.

Le dépôt légal des imprimés a pris en charge en 2009 environ 65.000 ouvrages, soit un chiffre similaire à l'offre commerciale d'*ebooks* en 2010 selon le MOTif. En considérant dans un premier temps une mise en place progressive du dépôt numérique, on peut raisonnablement estimer un premier effort de collecte d'*ebooks* entre 10 et 30.000 ouvrages par an¹⁴⁰. En considérant un poids moyen par *ebook* de 5 à 10 Mo au plus, le poids des données collectées atteindrait 50 à 300 Go par an.

Le traitement de cette masse d'*ebooks* progressivement recueillis pourrait se répartir, à l'instar du traitement du papier, sur tous les services du dépôt légal, avec un impact réduit sur l'organigramme. La supervision de l'étape de dépôt (dont la vérification du format et des métadonnées) pourra être dévolue au service des entrées ; aux équipes du catalogage reviendra l'information bibliographique, son enrichissement éventuel et sa mise en forme, puis son versement au catalogue général ; l'équipe de l'Extranet travaillera comme aujourd'hui sur le développement, le suivi et la maintenance du service, ainsi que le dialogue avec les déposants. Le problème d'une telle organisation calquée sur l'existant se situe essentiellement sur le point de la formation du personnel : elle nécessite des compétences qui ne sont pas forcément partagées par toutes les équipes, et qui demanderont un certain temps avant d'être acquises par l'intermédiaire de formations spécifiques. Il serait de toute façon préférable que seuls certains membres de chaque équipe prennent en charge le traitement des documents numériques.

L'autre solution, peut-être plus indiquée, au moins pour les premiers temps d'expérimentation, consisterait en un resserrement du traitement des *ebooks* ainsi collectés sur une équipe particulière, dotée des compétences nécessaires. Cette équipe, construite à partir des responsables de l'Extranet, gérerait alors toutes les étapes de traitement, dès l'entrée du document par l'Extranet, sa vérification, puis les procédures de signalement, et éventuellement de conservation. Cela aurait comme avantage, hors de la question inévitable des compétences, de cerner tout de suite les modifications à faire sur la procédure de dépôt, car une vue condensée sur la chaîne de traitement serait plus facile.

Se pose en tout cas la question de la fin de la chaîne, une question qui existe également pour le dépôt légal du Web. En théorie, le DDL prend en charge les

plusieurs experts de préservation. On devra également compter avec un supplément de bande passante mise à disposition du service.

140. En comparaison, il y a aujourd'hui environ 28.000 livres sous droits dans Gallica.

documents jusqu'au catalogage, puis passe la main à la DCO pour les opérations de conservation, de diffusion, de valorisation et de médiation. En ce qui concerne les *ebooks*, dans un premier temps, les opérations de versement pour consultation ou pour préservation seraient peut-être à effectuer directement par l'équipe de dépôt légal, au moins dans le cadre de l'expérimentation, et dans un but de gain de ressources. Cependant, la DCO ne doit pas perdre de vue ces collections — dont la valorisation lui revient — et doit être bien présente dans les réflexions qui les entourent. Faudrait-il alors basculer certaines ressources d'un département sur l'autre ? Sachant que sur le long terme, les départements des collections seront de toute façon amenés à prendre en charge de manière croissante ces documents numériques. Cette interrogation met en jeu des responsabilités entre les départements qu'il s'agira de bien évaluer, car elle conditionne toute la vision des collections numériques au sein de la BnF.

L'établissement pose dans ses priorités de gestion du personnel la formation. Elle se révélerait ici primordiale, en raison des aspects méconnus et techniques de certaines des étapes du traitement. Ainsi, une initiation au livre numérique dans toutes ses dimensions pourrait être la bienvenue pour permettre au personnel concerné d'acquérir des compétences en ce domaine. La question de ces compétences se poserait évidemment aussi au moment des recrutements. En effet, plus largement, l'*ebook* et sa gestion pose des questions d'organisation qui interrogent directement le métier de bibliothécaire. La chaîne du numérique est difficile à calquer sur celle de l'imprimé et ne met pas en jeu les mêmes qualifications¹⁴¹.

III. AVANTAGES ET INCONVENIENTS

A. Avantages

Cette solution offre plusieurs avantages pour pallier des déficits actuels.

Tout d'abord, elle permet de viser les *ebooks* payants, qui sont aujourd'hui difficiles à atteindre, en l'absence du décret d'application dont nous avons déjà parlé. Ces *ebooks* payants sont en effet accessibles par des processus divers d'identification et de téléchargement qui les rendent pour le moment quasi hors de portée du DLN sans difficultés techniques et juridiques. Un dépôt sur l'Extranet s'assimilerait à ce qui se fait déjà pour les ouvrages imprimés : le déposant remet directement à la BnF l'ouvrage qui fait l'objet du dépôt légal.

Ensuite, cette procédure – et c'est là sans doute son vrai avantage – donne accès à une véritable possibilité de référencement de l'ouvrage. Entrant dans une filière similaire à celle qu'empruntent actuellement les ouvrages imprimés, le livre numérique peut passer par une phase de catalogage, à partir des métadonnées éventuellement récupérées autour du fichier. Son signalement dans le catalogue général de la BnF et dans Gallica¹⁴² est rendu plus facile par son statut en quelque sorte individualisé : il bénéficie d'une notice propre, plus ou moins indépendante de sa version sous forme papier si celle-ci existe. À partir de cette notice, le lecteur pourra accéder soit à un lien vers l'interface de lecture du fichier dans le cas d'une consultation en Rez-de-Jardin, ou bien vers Gallica qui ouvre sur l'ouvrage ou sur un extrait et un renvoi vers un e-distributeur.

141. Ces aspects métiers sont notamment abordés par E. BERMÈS, La conservation des documents numériques, *op. cit.* p. 223.

142. Si ces *ebooks* sont accessibles à la consultation dans Gallica, l'avantage en termes de référencement est que le lien entre la bibliothèque numérique et le catalogue général est déjà mis en place.

Par ailleurs, la réflexion autour des déposants particulièrement concernés par ce mode de dépôt pourrait aboutir à une efficacité redoublée de la collecte d'*ebooks*. En effet, en ciblant les e-distributeurs ou les éditeurs de livres numériques, la bibliothèque pourrait approcher une quasi exhaustivité dans le domaine du dépôt des *ebooks*, comme pour les ouvrages papier aujourd'hui.

Se pose toujours la question de l'auto-édition, mais il semble qu'en portant une attention soutenue aux plateformes organisant cet aspect, on pourrait atteindre une partie importante des ouvrages produits par cette filière. Sur certains sites, on trouve clairement la mention du dépôt légal, et la précision de l'existence d'un dépôt des *ebooks* sur ces sites, voire un lien vers l'Extranet éditeurs, donnerait une publicité sans aucun doute satisfaisante à cette procédure. L'attention à porter à l'auto-édition est cruciale : en raison de son faible coût, on peut penser qu'elle ira croissant, et qu'elle concerne avant tout des ouvrages qui ne sont disponibles qu'en format numérique. Non tirés et en l'absence d'un dépôt légal des *ebooks* connu et clairement affirmé, ces *ebooks* peuvent facilement échapper à la bibliothèque, qui n'en garderait donc aucune trace.

Enfin, l'utilisation d'une interface telle que l'Extranet éditeurs revêt des aspects pratiques sur le plan technique. Au préalable, il faudrait mener une étude des formats disponibles et de ceux qui semblent le plus adaptés à la conservation pérenne et à la mise en place d'une consultation aisée. À partir des résultats de cette étude, le formulaire de dépôt pourra proposer quelques formats sélectionnés pour le fichier déposé. Cette limitation simplifierait les traitements en aval, dont la mise à jour des fichiers et leur éventuelle conversion au sein de SPAR.

B. Inconvénients

En revanche, il est des aspects qui manquent à ce scénario, et qu'il ne faut pas sous-estimer, à commencer par la perte relative de l'environnement d'usage dans lequel peut se placer l'*ebook*. La documentation du cadre éditorial est rendue impossible dans le cas où le dépôt n'est celui que d'un fichier contenant l'ouvrage et rien que lui. Ainsi, le site web, la place de l'ouvrage au sein du catalogue de l'éditeur ou du e-distributeur, voire tous les usages plus ou moins interactifs (mini-sites) qui pourraient être proposés autour du fichier, ne seraient pas concernés par ce dépôt, ce qui est évidemment dommage. Certes, on peut penser que ces usages seront peut-être enregistrés par le biais du dépôt légal du web, mais il est difficile de s'en assurer.

Des difficultés techniques se posent également, comme la question des métadonnées. Si le format ONIX est très utilisé aujourd'hui par les libraires et les e-distributeurs, ce n'est pas le cas de tous les éditeurs. On pense notamment à l'auto-édition, qui n'a pas forcément usage d'autant d'éléments que ceux que propose ONIX. La BnF peut cependant se permettre d'encourager l'utilisation de certains outils, et les recommandations de plusieurs acteurs du monde de l'édition numérique placent certains formats (EPUB, ONIX) dans les recommandations qui permettraient de considérer le futur de l'*ebook*, bibliothéconomique comme éditorial, d'un œil plus serein. Les DRM seront un autre point problématique et ce scénario ne peut qu'induire une négociation renforcée avec les producteurs, afin d'obtenir de leur part une utilisation la plus souple possible des systèmes de gestion de droits attachés aux fichiers déposés.

Enfin, il reste une incertitude sur la volumétrie des *ebooks* concernés et sur le nombre d'éditeurs susceptibles d'être touchés et intéressés par un tel service. À côté des nés numériques qui doivent mériter une attention particulière, il faut prêter attention aux ouvrages qui pourront, à partir d'un moment donné, être déposés par ce biais au titre du dépôt légal de substitution. Le basculement de ces dépôts du papier au numérique doit être attentivement surveillé et pris en compte, même s'il s'agit a priori d'un ensemble de

documents facile à appréhender. Le manque de statistiques aujourd'hui disponibles entache d'un certain flou les raisonnements que l'on peut avoir sur le sujet.

Ces incertitudes sur la quantité de documents à prendre en compte pèsent sur le dernier point problématique de cette solution, à savoir l'organisation au sein du DDL, et plus généralement de la BnF. Une organisation trop lourde et trop calquée sur l'imprimé risquerait de manquer de la flexibilité nécessaire, alors que la constitution d'une équipe à part pourrait peser sur la situation des *ebooks* dans les collections de la bibliothèque.

Chapitre 6 : Un dépôt par le moissonnage du Web

Le dépôt légal des *ebooks* pourrait également se faire par le biais du dépôt légal du Web. Le raisonnement est un peu différent, par rapport à une solution Extranet, et ne peut avoir de parallèle réel avec le papier. Avec le dépôt légal du Web, on est à la fois dans une logique d'échantillonnage (toutes les mises à jour de tous les documents ne peuvent être capturées), mais aussi dans une visée de collecte plus large (le fichier de l'*ebook* ne représente pas le niveau de granularité privilégié). Cependant, c'est bien la même volonté de représentativité d'une production éditoriale dont il est question.

I. LE DEPOT LEGAL DU WEB

Le dépôt légal de l'Internet existe à la BnF depuis 2006, nous en avons expliqué le contexte législatif et institutionnel dans les grandes lignes (chapitre 3). En mettant en place ce type de dépôt, il est question d'insister sur la valeur patrimoniale et éditoriale de la production publiée sur Internet. Les *ebooks* étant l'un des axes de cette production, il est logique d'envisager leur collecte par ce biais.

Le dépôt légal numérique a été confié à un service particulier de la BnF, intégré au DDL, qui comptait en 2009 9,5 ETP répartis entre le DLN et le DSI, mais aussi 85 correspondants pour le dépôt légal du Web répartis dans toute la bibliothèque¹⁴³.

Malgré le flou juridique que crée l'absence du décret d'application de la loi DADVSI, le DDL organise la collecte des données numériques à titre expérimental. Il s'agit de ne pas prendre de retard sur la patrimonialisation d'Internet, et de ne pas passer à côté des « incunables du Web ». Cette entreprise se structure autour de deux axes principaux :

- des collectes automatiques du domaine français tel que précédemment défini¹⁴⁴, peu profondes, effectuées une fois par an, qui ont pour objectif de réaliser un cliché superficiel du Web français à un instant donné ;
- des collectes ciblées, plus profondes pour une image plus fine, sur certains sites choisis suivant les sélections des correspondants du DLN dans les départements ou bien selon des projets particuliers¹⁴⁵.

Ces moissonnages du web se font sans une participation active du déposant. Le DLN prend en main la collecte, grâce au robot Heritrix¹⁴⁶ qui fonctionne comme un internaute automatique : partant d'une liste d'URL qui servent de points d'origine (de « graines »), il suit les liens qu'il trouve sur les pages et copie les éléments qu'il y détecte¹⁴⁷. Cela permet de disposer d'un instantané cohérent du Web et de constituer un état de l'Internet national à un moment donné. La BnF avait depuis 2004 un accord avec

143. Les correspondants des départements proposent au DLN des listes de sites à collecter, les « paniers ».

144. Voir le chapitre 3 : il s'agit actuellement des sites Web enregistrés en .fr et affiliés (.re), ainsi que des sites édités par des personnes physiques ou morales domiciliées en France, quelle que soit leur extension (.com, .net, .org, etc.).

145. C'est dans cette optique qu'ont été réalisées les collectes ayant pour thèmes le Web militant, les élections européennes de juin 2009 ou bien les journaux intimes.

146. <<http://crawler.archive.org>>

147. Voir G. ILLIEN, C. OURY. Quelle politique documentaire pour l'archivage des sites Internet ? *Les collections électroniques...*, op. cit, p. 157-178.

la fondation Internet Archive¹⁴⁸ pour les collectes annuelles du Web ; ces collectes sont totalement internalisées depuis 2010. La bibliothèque a également acquis un échantillon des archives du Web français depuis 1996 auprès de cette fondation, et entretient avec elle un accord de partenariat et de recherche.

Ces collections représentaient en 2009 189 To de données, soit 13,67 milliards d'URL. Il s'agit donc de volumes sans commune mesure avec les flux de l'imprimé. Leur consultation s'effectue dans les salles du niveau Recherche depuis 2008 grâce à un logiciel d'indexation fournissant une interface d'interrogation, la Wayback Machine¹⁴⁹, créée par Internet Archive. La conservation de ces archives du Web se fera dans SPAR, la filière concernée étant en cours d'instruction, pour une perspective de versement dans la seconde moitié de 2011¹⁵⁰.

Gildas Illien présente dans un article de 2008¹⁵¹ les différents outils utilisés pour l'archivage du Web. Ces logiciels sont libres et développés de manière collaborative par Internet Archive et les bibliothèques membres du consortium IIPC¹⁵². Toutes les étapes de traitement des archives de l'Internet font l'objet d'une automatisation large. La masse des données à traiter et l'urgence due à leur caractère éphémère exigent un gain de temps maximum et le moins d'opérations manuelles possible. Le caractère automatique de la collecte est inscrit dans la loi¹⁵³, même si celle-ci précise que d'autres modalités pourront être mises en œuvre si la méthode automatique ne permettait pas de capturer certains sites, en accord avec les éditeurs.

II. QUELLES POSSIBILITES POUR LE DEPOT LEGAL DES EBOOKS ?

Le dépôt légal de l'Internet a été institué dans l'optique de collecter le Web dans son ensemble, plutôt que ses parties¹⁵⁴. Néanmoins, au vu de l'évolution du Web et de ses possibilités, l'inscription de la question des *ebooks* dans les objectifs du DLN paraît tout à fait judicieuse, au vu de la place croissante qu'ils sont probablement appelés à occuper dans la production numérique française.

En 2009 une expérience de collecte automatisée d'*ebooks* a été conduite sur un ensemble de plateformes de diffusion de livres numériques. Elle avait pour objet de préciser les possibilités, les limites et les futures étapes possibles d'une capture automatique d'*ebooks* par voie de moissonnage. En l'absence du décret nécessaire pour la collecte de l'offre payante, seuls les livres numériques gratuits étaient concernés¹⁵⁵. Une liste de sites à analyser a d'abord été définie, au vu du paysage commercial d'une part et de l'état des archives disponibles d'autre part¹⁵⁶, sachant qu'aucune collecte

148. Site officiel : <<http://www.archive.org/>>

149. <<http://archive-access.sourceforge.net/projects/wayback>>

150. Concernant SPAR, voir le chapitre 5.

151. G. ILLIEN. Le dépôt légal de l'internet en pratique : les moissonneurs du Web. Paris : BBF, 2008. *BBF*, vol. 53, n° 6. Disponible à l'adresse : <<http://bbf.enssib.fr/consulter/bbf-2008-06-0020-004>> (consulté le 7 décembre 2010).

152. International Internet Preservation Consortium. Site officiel : <<http://netpreserve.org/about/index.php>>

153. Dans la loi DADVSI, l'article L. 132-2-1 parle de « collecte selon des procédures automatiques ».

154. Comme le fait remarquer à propos des périodiques en ligne Maryline Devidal dans son projet professionnel personnel *Stratégies pour un dépôt légal des périodiques en ligne – Dossier d'aide à la décision* (FIBE, ENSSIB, 2010).

155. Selon une définition qui correspondait à celle que nous avons vue dans notre chapitre 1 : des fichiers, comparables à des monographies.

156. Étaient notamment présents dans ces archives les sites Cyberlibris, Ebookslibresetgratuits, Eyrolles, Mobipocket, Numilog, Publie.net. Mais la plupart des captures antérieures de ces sites ne comportaient aucun *ebook*, ou bien des extraits seulement — à l'exception faite de Publie.net, qui faisait l'objet d'un partenariat particulier.

spécifique centrée sur les *ebooks* n'avait été entreprise jusque là. La collecte de test sur ces sites représentatifs¹⁵⁷ a été lancée en mars 2009, durant 26 jours¹⁵⁸. Notre réflexion se base en partie sur cette expérience.

A. La collecte

1. Les collectes larges

Il s'agit d'inclure la collecte des *ebooks* dans le processus de moissonnage du Web tel qu'il existe aujourd'hui. D'ailleurs, la BnF a déjà collecté des livres numériques dans le cadre des précédentes collectes larges, sans que l'accent soit nécessairement mis sur ce type de documents. Lors de ces opérations, la capture des sites se fait de manière très étendue et, comme nous l'avons souligné, l'exhaustivité n'est pas l'objectif. Mais il s'agit de l'outil qui permet la prise en compte la plus complète de la diversité des *ebooks* sur Internet. La collecte large permet de récupérer sans effort de sélection particulier tout l'environnement de production et de diffusion des *ebooks*, voire certains d'entre eux qui échapperaient à une prospection plus ciblée.

Par ailleurs, pour les besoins de notre étude, nous avons défini (chapitre 1) l'*ebook* comme un objet répondant à quelques critères précis ; la collecte large a cependant l'avantage de retenir également un certain nombre d'autres formes de publications numériques, dans l'attente d'une définition partagée par tous les acteurs du secteur. Sont ainsi capturés les sites gouvernementaux (rapports officiels), les sites d'archives ouvertes des universités, les pages personnelles d'écrivains amateurs qui ne passent pas par des services d'auto-édition, etc.

Ces collectes larges ont une vocation avant tout exploratoire et ne capturent donc les sites que superficiellement. Leur limite est d'ordre qualitatif¹⁵⁹ : c'est l'URL du site qui héberge la ressource (en l'occurrence l'*ebook*) qui sert de point de départ à la collecte et non l'adresse de cette ressource même. Ainsi, il est possible que les robots des collectes larges ne pénètrent pas assez profondément dans le site pour atteindre l'*ebook*. Cet inconvénient, conséquent pour certains des sites qui nous intéressent, peut être avantageusement contrebalancé par des collectes plus ciblées. Cependant, pour des sites à l'architecture simple, ce mode de collecte suffirait à atteindre des documents comme les *ebooks*.

Le modèle technique de conception et de diffusion des sites a un fort impact sur leur moissonnage. L'efficacité des robots moissonneurs peut être limitée par la structure du site, mais peut également buter sur des barrières telles que les formulaires ou les contenus accessibles derrière un mot de passe ou une identification. Or nous avons vu que la plupart des plateformes de diffusion d'*ebooks* conditionnaient l'accès à leurs ouvrages à une identification, ce qui pour le moment les rend difficiles ou impossibles à collecter. L'exemple du portail de revues CAIRN, accessible aux robots de la BnF grâce à la reconnaissance de leur adresse IP, est cependant un exemple encourageant pour le futur¹⁶⁰. Une procédure de e-dépôt à l'unité peut éventuellement être mise en œuvre dans les cas où l'architecture d'un site sélectionné rend impossible la collecte.

157. Les sites collectés sont indiqués par un astérisque dans la liste que nous proposons en annexe 5.

158. Au total, au moins 1.623 livres numériques gratuits ont été capturés lors de cette collecte.

159. Les collectes larges ne peuvent garantir la complétude d'un site dépassant les 10.000 URL (G. ILLIEN, C. OURY. Quelle politique documentaire..., *op. cit.*, p. 172) ; en comparaison, on considérera par exemple les 47.900 résultats du site Numilog dans son entier selon une recherche Google au 2 décembre 2010. Lors de l'expérimentation de 2009, ce site comptait dans les archives 8.100 URL, avec uniquement des extraits au format PDF (1.200).

160. *A contrario*, l'exemple de Publie.net appelle à une attention soutenue : ce site a d'abord collaboré avec le dépôt légal, en donnant un accès à ses collections, mais l'architecture en a récemment été changée, et les robots ne peuvent plus accéder aux fichiers.

2. Les collectes ciblées ou courantes

Ce volet de l'opération se placerait dans le cadre des collectes courantes, qui concernent aujourd'hui des sites que la BnF a vocation à archiver dans la durée et dans la profondeur. L'idéal serait de continuer à travailler avec la DCO afin de dresser une liste de sites cibles, comme c'est le cas actuellement pour certains sites signalés par les départements dans leurs « paniers ». Définir une liste de sites prioritaires dans la récolte permet de circonscrire une production éditoriale représentative, après une étude approfondie de l'offre dans le domaine des *ebooks*. Pour cibler ces sites, on se concentrerait sur ceux qui mettent en place des usages spécifiques au livre numérique tels qu'on les a évoqués plus haut. Il s'agirait avant tout des plateformes de diffusion des *ebooks*, payants ou gratuits, quelle que soit l'étendue de leur offre. L'objectif est d'obtenir un échantillon diversifié de l'offre sur le marché français, sans les restrictions que peuvent poser les collectes larges (profondeur) ou la solution Extranet (formats notamment).

Selon des critères d'offre ou de tendances éditoriales (dont l'auto-édition), on peut proposer une première liste¹⁶¹, à partir notamment de celle déterminée pour l'expérimentation de 2009¹⁶². Cette liste, qui demande sans doute à être enrichie, prend pour critères :

- l'étendue de l'offre proposée : des sites comme Numilog permettraient d'obtenir un maximum d'*ebooks* ;
- le contenu proposé : un site d'éditeur comme celui de Gallimard renvoie pour l'achat à des distributeurs, mais propose des rubriques autour de ses ouvrages voire des mini-sites intéressants ;
- la proposition effective d'*ebooks* sur le site : les sites d'éditeurs ne proposant pas l'accès à des livres numériques n'ont pas été retenus ;
- la qualité de l'offre proposée : des sites d'édition contemporaine ou d'auto-édition ont été retenus, malgré la faiblesse relative de leur offre ;
- un partenariat déjà existant avec la BnF, qui assure déjà d'un dialogue avec la BnF et d'un intérêt pour le sujet.

Il sera bien entendu nécessaire de mettre régulièrement cette liste à jour, pour englober de nouveaux sites et éviter d'en collecter qui ne sont plus en activité. Par ailleurs, il convient de réfléchir à la périodicité de la collecte de ces sites, notamment en fonction de l'importance de la plateforme et du nombre d'*ebooks* qu'elle met régulièrement à disposition des internautes. Les correspondants des départements qui sélectionnent les sites doivent en effet choisir les paramètres de la récolte pour chacune des URL, dont la profondeur¹⁶³ et la fréquence¹⁶⁴. Ici, une profondeur maximale serait impérative, pour être sûr d'atteindre le fichier numérique de l'*ebook* lui-même. Quant à la fréquence, elle reste à définir selon les plateformes et leur taux de renouvellement ; actuellement, au vu de l'offre encore assez restreinte, il est probable que les ouvrages

161. Voir cette liste en annexe 5. La question de l'adresse de base donnée en graine au robot mérite d'être posée : moissonnera-t-on l'intégralité du site diffuseur ou bien seule la partie qui donne accès aux *ebooks* ? Le choix reviendra en ce domaine aux correspondants des départements.

162. Les sources étaient alors la liste des e-distributeurs partenaires de Gallica 2, les éditeurs cités dans le rapport Patino, et les éditeurs et diffuseurs ressortant lors d'une recherche sur Google et dans la veille professionnelle.

163. Domaine, hôte, page et son contexte, page seule ou chemin.

164. Une fois par semaine, une fois par mois, tous les 2 mois, capture unique.

mis à disposition restent sur le site assez longtemps avant d'en disparaître, ce qui permettrait d'espacer le rythme de collecte.

Enfin se pose la question de la relation entre les livres numériques et les ouvrages imprimés. Il conviendrait peut-être sur ce point de fixer une ligne générale afin d'assurer une certaine harmonie de la collecte : les sélectionneurs peuvent d'une part favoriser la continuité des collections, ce qui induit de collecter en numérique ce qui est déjà présent en papier dans les collections, selon le titre, l'auteur, l'éditeur. D'autre part, il peut s'agir de mettre en avant les publications nées numériques et inédites sous forme imprimée. Cette seconde solution encourage davantage la complémentarité que la liaison directe entre l'imprimé et le numérique, mais met en valeur la spécificité de ce dernier, tout en permettant une certaine économie de moyens (pas de doublons papier/numérique).

B. Le traitement : référencement, conservation, consultation

La perspective d'un traitement particulier des *ebooks* au sein des archives du Web ne va pas de soi. Les archives du Web sont conservées comme un tout au sein de SPAR, selon une filière qui leur est spécifique (DL_AUTO), actuellement en cours d'instruction. Elles sont stockées dans les formats containers (W)ARC¹⁶⁵. En termes de conservation, les livres numériques se placent au même rang que les autres objets du Web collectés. Le dépôt légal du Web capture ainsi des *ebooks* depuis cinq ans, sans les distinguer du reste de la collection. Cette distinction est encore malaisée à faire. Elle ne se justifie d'ailleurs que si l'on souhaite mettre en avant la spécificité de ce sous-ensemble des archives du Web ; principalement en raison de la question des formats et des DRM, ces fichiers se distinguent par une certaine fragilité (accès parfois non pérenne) et un statut particulier par rapport aux collections imprimées (continuité des collections). Ils constituent donc une certaine production éditoriale particulièrement intéressante qui justifierait une conservation et un traitement en partie distincts du reste des archives du Web¹⁶⁶.

Il faut garder à l'esprit la spécificité des collections numériques et la nécessité de ne pas vouloir à tout prix leur imposer un cadre bibliothéconomique traditionnel. Ainsi, le modèle de production du dépôt légal des sites Web ne prévoit pas leur catalogage à l'échelle du site, et encore moins à celle des fichiers, essentiellement aujourd'hui pour des raisons économiques. En ce qui concerne les *ebooks*, la même logique est de mise, car ils sont appréciés comme un sous-ensemble du Web, intégré à celui-ci. Pourtant, il serait intéressant de pouvoir exploiter les métadonnées décrivant les *ebooks* qui entrent par le biais des collectes dans les collections numériques, à l'instar de ce que nous avons vu dans la solution Extranet. Ceci permettrait notamment de retrouver plus facilement ces ouvrages au sein des archives du Web (voir plus loin).

Il paraît cependant incontournable de passer par un traitement automatique de ces métadonnées. Certaines institutions¹⁶⁷ se sont essayées à envisager une extraction de

165. Pour le format de compression ARC (ARChive), voir <<http://www.archive.org/web/researcher/ArcFileFormat.php>> ; et pour le format WARC (Web ARChive), qui est une extension du format ARC et qui est désormais une norme ISO, voir <<http://bibnum.bnf.fr/WARC>> (consulté au 19 décembre 2010).

166. Et en partie seulement car tout ce qui fait la difficulté du sujet est là : l'intérêt de récolter les *ebooks* dans le cadre des archives du Web est de capturer en même temps les ouvrages et leur contexte éditorial. Il ne faut donc surtout pas purement et simplement extraire les *ebooks* des archives du Web après récolte, sans tenir compte du reste du contexte archivé, mais allier les deux aspects.

167. On pense notamment à the Humanities Advanced Technology and Information Institute (HATII) de l'université de Glasgow, qui a réfléchi à la comparaison pour la production et l'extraction des métadonnées entre une méthode manuelle et une méthode automatique, en circonscrivant la réflexion aux objets au format PDF dans une collection numérique (hors archives du

métadonnées manuelle, mais vu l'ampleur des archives du Web de la BnF, cette solution est hors de propos¹⁶⁸. L'extraction des métadonnées peut être compliquée par le fait qu'elles peuvent être ou bien incluses dans le document ressource lui-même, ou bien autonomes. Il ne suffirait donc pas de cibler les fichiers d'*ebooks* puis d'en extraire automatiquement les informations descriptives. Plusieurs solutions pour cette extraction peuvent être considérées, selon le format d'origine des métadonnées et les informations concernées. D'autres bibliothèques, comme celle du Congrès¹⁶⁹ ou celle du Canada¹⁷⁰, réfléchissent à ce problème de l'extraction des métadonnées. Il conviendrait bien sûr d'étudier ces projets afin de trouver ce qui s'adapterait le mieux au traitement des collections de la BnF. Si l'on arrivait à traiter ces métadonnées, on pourrait envisager un lien entre l'ouvrage récolté et une notice tirée de ces informations dans le catalogue général.

Le référencement de l'ouvrage grâce à l'exploitation des métadonnées n'est pas la seule manière de rendre les *ebooks* plus visibles. Pour le moment, afin d'accéder aux *ebooks* entrés par le dépôt légal du Web, le lecteur doit passer par la Wayback Machine, qui permet la recherche par URL dans les archives du Web. Ce mode de recherche n'est guère satisfaisant dans le cadre du traitement d'une collection d'*ebooks*, puisque le lecteur devrait connaître l'URL exacte de l'ouvrage qu'il recherche. Le développement de la recherche par mots, et donc de l'indexation plein texte des contenus apportera à la consultation de ces collections un indéniable avantage. Pour cela, la BnF utilise le logiciel NutchWAX¹⁷¹, qui fonctionne encore d'une manière partielle et expérimentale (sur environ 10 % du fonds). Dans l'optique *ebooks*, cette option de recherche paraît aujourd'hui indispensable, surtout si l'on considère l'absence de traitement bibliographique de ces données, ce qui exclut encore une recherche dans le catalogue. Il faut réfléchir à cette recherche pour éviter trop de bruit autour d'un titre ; on pourrait utiliser certaines métadonnées comme des mots-clés afin de la rendre plus pertinente. L'indexation et le référencement seraient à considérer comme un tout pour permettre au livre numérique d'être plus accessible.

Afin d'exploiter et de conserver au mieux les *ebooks* collectés, il faudrait d'abord pouvoir les repérer dans les archives du Web, ce qui passe notamment par l'identification du format des fichiers. Selon la typologie des formats au sein de SPAR que nous avons évoqué plus haut (chapitre 4), les formats qui sont aujourd'hui conservés dans le système sont simplement identifiés (voire pour une part seulement stockés), car la mise en place du système n'est pas encore totalement achevée, et que les possibilités logicielles sont encore limitées dans ce domaine. Pour notre sujet, il est important de noter que les méthodes d'archivage en ligne sont encore relativement primaires par rapport aux capacités de l'édition en ligne¹⁷².

Web). La conclusion se prononce nettement en faveur de processus automatisés. Voir : <http://www.digitalpreservationeurope.eu/publications/briefs/11_FRENCH.pdf> (consulté le 27 décembre 2010).

168. Que l'on pense aux proportions : 66.000 livres sont entrés au dépôt légal en 2009, contre 1.270 millions de fichiers dans le cadre du dépôt de l'Internet.

169. La Library of Congress enrichit depuis 2005 ses notices en MARC grâce à une extraction automatique de la table des matières des ouvrages. Voir l'article *Machine-generated Contents Notes* sur le site du Bibliographic Enrichment Advisory Team : <<http://www.loc.gov/catdir/beat/mg505.html>> (consulté le 2 janvier 2010). Cette équipe explore également les possibilités du format ONIX.

170. Bibliothèque et Archives du Canada (BAC) utilisent quant à elles le logiciel Interwoven MetaTagger 4.1.0 pour extraire les informations des paquets de données numériques et les inclure dans une notice METS. Voir <<http://www.collectionscanada.gc.ca/digital-initiatives/012018-4000.01-f.html>> (consulté le 3 janvier 2010).

171. De Nutch (moteur de recherche open source) et de Web Archive eXtension. Voir : <<http://archive-access.sourceforge.net/projects/nutch/>>.

172. G. ILLIEN. Le dépôt légal de l'internet en pratique, *op. cit.*

Ainsi, il n'est pour le moment pas possible de trier automatiquement les fichiers qui ont été capturés sur les sites de diffusion d'*ebooks*. La seule information sur la nature des documents dont on dispose dans les rapports du robot moissonneur est leur « type MIME », c'est-à-dire l'identifiant de leur format tel que déclaré par les éditeurs de sites. Dans le cas des *ebooks*, quelques-uns de ces types MIME peuvent être exploités sans ambiguïté¹⁷³, mais cette information est souvent inutilisable, car elle n'est pas propre aux livres numériques spécifiquement. C'est le cas notamment des fichiers PDF, MP3 ou WMA. Les fichiers au format PDF par exemple sont particulièrement répandus sur le Web, et sont donc collectés largement par le dépôt légal du Web ; ils comprennent des *ebooks* comme d'autres documents : plus de 7 millions de documents en PDF ont été capturés par la collecte large de 2008, réalisée par Internet Archive. Parmi ces millions de PDF, tous ne sont bien sûr pas des *ebooks*, mais le tri reste impossible à faire.

Des outils¹⁷⁴ existent pourtant qui permettent d'identifier et de valider automatiquement les formats de fichiers et les informations qui leur sont reliées à leur entrée dans une collection numérique¹⁷⁵. On peut citer les logiciels DROID¹⁷⁶, qui identifie les formats des fichiers par lots, ou Jhove¹⁷⁷, qui analyse un fichier pour valider son intégrité et en extraire (XML) les métadonnées qui le décrivent¹⁷⁸. Cependant, l'échelle des performances de ces outils ne rend pas encore possible à moindre frais l'identification et la validation systématiques de tous les fichiers issus des collectes, ce qui autoriserait un traitement réellement satisfaisant dans le cadre du livre numérique. Ces limites dans le traitement des formats de fichiers posent évidemment plus largement la question de la conservation dans SPAR, et leur remise en question évoluera avec la mise en place du système.

La consultation des résultats des collectes suit la même restriction que celles appliquée aux imprimés acquis par la BnF par le dépôt légal, c'est-à-dire qu'elle est limitée par la loi aux chercheurs accrédités, et se fait donc uniquement dans les espaces de consultation qui leur sont dédiées (Rez-de-Jardin, salles de lecture de Richelieu, etc.).

Nous n'envisageons ici que la consultation des *ebooks* sur les postes donnant accès aux archives du Web. Les livres numériques capturés sur des plateformes hors Web et accessibles depuis des liseuses sont pour le moment l'une des difficultés juridiques ; par ailleurs, la mise à disposition de liseuses, comme nous l'avons précisé (chapitre 4), demanderait des critères délicats à déterminer. Cependant, comme pour le scénario précédent, l'environnement de consultation publique de la BnF ne propose pas vraiment des applications de lecture appropriées et satisfaisantes pour la consultation de ces *ebooks*, même présents dans les archives du Web. L'interface lecteur ne permet pas d'installer de tels logiciels, il faudra donc que la bibliothèque mette ces outils en

173. Il en est ainsi pour : chemical/x-pdb (fichiers PDB), application/vnd.palm (fichiers PRC), application/x-mobipocket-ebook (fichiers Mobipocket), ou application/epub+zip (fichier EPUB). On trouvera une liste de types MIME à cette adresse : <<http://www.iana.org/assignments/media-types/>> (consulté le 26 décembre 2010). Attention, il arrive que le type MIME utilisé pour décrire les *ebooks* ne soit pas approprié, et donc reste inutilisable dans notre optique.

174. Voir la note sur les répertoires de formats d'E. Bermès sur son blog, du 6 janvier 2006, disponible à l'adresse <<http://www.figoblog.org/document1127.php>> (consultée le 25 décembre 2010).

175. On pense ici précisément aux collectes contemporaines, mais il ne faut pas perdre de vue qu'un même traitement devra être opéré sur les collectes déjà archivées, lorsque les outils le permettront.

176. DROID accède pour ses identifications au dépôt PRONOM, développé par les archives nationales anglaises (<<http://labs.nationalarchives.gov.uk/wordpress/index.php/2010/10/linked-data-and-pronom/>>). Site : <<http://droid.sourceforge.net/>>.

177. Site officiel : <<http://hul.harvard.edu/jhove/>>.

178. E. BERMÈS. La conservation des documents numériques, *op. cit.*, p. 221.

place¹⁷⁹. En revanche, la lecture en ligne, si le robot a pu capturer tout ce qu'elle implique, devrait rester possible directement depuis la Wayback Machine.

C. Coûts et moyens

Ce mode de collecte peut facilement se greffer sur le dispositif existant, et c'est là tout son avantage en termes de coûts et de moyens.

En ce qui concerne la collecte et la conservation de ces documents, les budgets s'expriment aujourd'hui en URL (en fichiers, ce qui nous intéresse au premier chef) et en octets (le stockage numérique est un enjeu crucial). Cette solution prendrait place dans le processus de conservation des archives du Web, économisant ainsi la création d'une filière spécifique. Il faudra mettre en perspective les différentes augmentations que requerra l'ajout des sites de diffusion d'*ebooks* aux collectes courantes actuelles (stockage, bande passante, etc.). Cependant, le poids de ces collectes serait assez marginal¹⁸⁰, et devrait pouvoir être absorbé par le flux courant du dépôt légal du Web.

Les collectes courantes présentent certes un rapport « temps de travail / nombre de données capturées » moins avantageux que celui des collectes larges¹⁸¹ ; cependant, il faut penser que ce sont elles qui permettraient les résultats les plus satisfaisants en matière de récolte d'*ebooks*. Si les collectes courantes sont plus onéreuses, c'est que la sélection des sites requiert le travail de bibliothécaires chargés de collection (A), appartenant aux départements des collections, qui seront les plus à même de choisir les paramètres de collecte appropriés. Il faudrait établir la part de budget qui serait allouée à la récolte des plateformes de livres numériques, par rapport au reste des « paniers » des correspondants.

La même question que pour la solution Extranet se pose, de manière tout aussi cruciale et délicate, sur la répartition des responsabilités et de la prise en charge du document entre le DLN et les départements des collections, avec la gestion des ressources que cela implique. L'expérience de 2009 avait souligné que la question des *ebooks* intéressait aussi bien le département Langues, Littérature et Arts (LLA) et le département de la Recherche bibliographique, que le DDL. Les moyens d'une telle opération seront donc à répartir entre les départements des collections (charge de travail supplémentaire possible pour les correspondants du DLN, parcours de valorisation ciblée des sites sélectionnés à mettre en place) et le DLN. Sachant que si actuellement les archives du Web sont gérées par le DLN, une partie de la gestion de ces collections devra sans doute passer tôt ou tard à la DCO, comme les collections analogiques et pour leur continuité.

Le service informatique sera logiquement touché par cette solution. Un effort particulier de développement devra être fourni sur la question des reconnaissances de formats (*ebooks* comme métadonnées) ; la recherche notamment des logiciels les plus appropriés et l'étude des solutions envisagées par d'autres bibliothèques occupera un certain temps de prospection, qui pourra éventuellement être partagé avec le DLN. Il est aussi nécessaire d'aborder le point de l'identification des robots sur les sites dont le contenu est protégé : si l'accord avec CAIRN montre la faisabilité d'une identification par IP, il faudra réfléchir à d'autres procédures afin de pouvoir s'adapter au mieux à

179. Voir le chapitre 4.

180. Ces documents sont pour la plupart des documents texte, dont le poids est sans doute secondaire par rapport par exemple aux efforts fournis pour mieux collecter la vidéo sur Internet.

181. G. ILLIEN, C. OURY. *Quelle politique documentaire...*, *op. cit.*, p. 176.

chaque cas (fourniture des mots de passe, accès direct au serveur, etc.). Par ailleurs, nous avons vu à quel point une réflexion d'ampleur était nécessaire pour l'adaptation de l'interface de consultation des archives du Web, ce qui est également du ressort du service informatique.

On pourra aussi approfondir la mise en place de contrôles qualité, afin de vérifier si les sites visés, notamment par les collectes ciblées, ont bien été capturés, à la bonne profondeur, et qu'ils sont accessibles. Ces contrôles qualité pourront à la fois permettre d'ajuster des paramètres tels celui de la profondeur ou de la fréquence, mais également vérifier si les formats et tous les fichiers des livres numériques ont bien été pris en compte et sont lisibles¹⁸². Leur automatisation serait souhaitable, au moins en partie.

Enfin, il faut estimer les paramètres d'entente avec les éditeurs des sites concernés. Certes, l'accord de ceux-ci n'est pas rendu impérativement nécessaire par la loi, puisque les textes précisent que la BnF peut passer outre leurs réticences pour collecter les contenus numériques. Un dialogue est néanmoins indispensable dans la perspective de la collecte des contenus protégés, et *a fortiori* des contenus payants lorsque le décret l'autorisera pleinement. Ce dialogue pourra être mené par le DLN, qui maîtrise exactement les tenants et les aboutissants pratiques d'une telle entente, en partenariat avec le service informatique¹⁸³. Les correspondants à la DCO pourront éventuellement participer à la spécification des interlocuteurs à privilégier, selon l'importance et les nécessités des sites et de leur offre.

Par ailleurs, le problème de la gestion des droits attachée aux fichiers capturés se pose de façon sans doute plus prégnante ici que dans la solution Extranet où l'on pouvait envisager une négociation en amont avec les éditeurs pour des dépôts sans DRM. Les fichiers collectés seraient a priori ceux mis à la disposition du public et donc verrouillés par le système de droits de l'offre publique. Ce point peut poser d'importantes difficultés pour la consultation, et le service informatique devra étudier les possibilités de contourner ces verrous, comme la loi l'y autorise. Mais cet aspect devrait encourager la BnF¹⁸⁴ à dialoguer avec les éditeurs numériques dans leur ensemble sur cette question.

III. AVANTAGES ET INCONVENIENTS

A. Avantages

L'expérience de 2009 a montré la faisabilité technique d'une collecte des livres numériques par robot, au titre du dépôt légal du Web. Elle a montré ce que cette modalité de dépôt avait d'économique : le surplus de ressources nécessaires, humaines notamment, est faible, puisque le principe de fonctionnement du dépôt légal du Web arrive à maturité. Ainsi, ce mode de collecte peut s'insérer sans difficultés majeures ni surcoût important dans un cadre juridique et un flux de production existants.

Cette économie en termes de ressources est d'autant plus intéressante à considérer en rapport avec le nombre important de fichiers que cette méthode permet de capturer. Ces modalités de collecte sont aussi intéressantes sur leur étendue (collectes larges) que sur leur profondeur (collectes ciblées), et permettraient donc d'arriver à une bonne représentativité du paysage éditorial de l'*ebook* français. Les questions importantes et

182. G. ILLIEN. Le dépôt légal de l'internet en pratique, *op. cit.*

183. Mais il convient de noter que des membres du service informatique sont déjà attachés au DLN et pourraient participer à la réflexion sur cette question.

184. Reste à préciser à quel niveau, sous quelle forme (groupe de travail transversal ?) et avec la participation de quels départements et services parmi ceux qui sont concernés.

déliçates à cerner de l'auto-édition et de la publication à la demande notamment sembleraient recevoir par cette méthode une réponse intéressante car relativement complète (dans le nombre d'ouvrages collectés comme dans leurs contextes de diffusion). Cette solution permet de ne pas s'attarder sur une définition exacte de ce qu'est un *ebook* dans un contexte de collecte, et donc d'en garder une vision très large, en maintenant ouvertes plusieurs options en attendant que ce point se définisse.

Un autre avantage de ce scénario le situe dans une logique patrimoniale. Les *ebooks* sont collectés dans leur contexte éditorial : page d'accueil, description de l'offre de services, liens vers les sites associés, animations, etc. Cela permet de mettre en avant la vraie spécificité du livre numérique que sont les nouveaux usages qui lui sont liés, ces usages « nés numériques », et de considérer les publications de l'Internet comme un tout cohérent. On pourrait ainsi proposer, sur l'exemple des précédentes exploitations des collections (journaux intimes, élections), un ou plusieurs parcours de valorisation autour du livre numérique ou d'un secteur éditorial en particulier, mettant en avant l'intérêt de la collecte de l'environnement de diffusion.

B. Inconvénients

Pour le moment, ce mode de collecte souffre particulièrement de l'absence du décret d'application l'ouvrant vers la capture des *ebooks* payants. Cependant, en l'attente de ce texte, il serait judicieux de développer les relations de confiance déjà tissées avec les éditeurs. À l'instar de CAIRN ou de Publie.net (jusque récemment), ceux-ci pourraient être encouragés à faciliter la collecte de leurs catalogues d'*ebooks* par le biais du dépôt légal automatique, en autorisant les robots de la BnF à parcourir leur site. Ceci est particulièrement valable pour les gros sites, à la structure complexe, qui ne manqueront pas de se développer au fur et à mesure que le marché du livre numérique s'étendra.

L'absence d'indexation et de catalogage des sites web rend difficile le référencement des *ebooks*, dans une optique bibliothéconomique de signalement et de valorisation. Le traitement approfondi des données délivrées par les fichiers numériques est coûteux, et ne pourrait donc être réservé qu'à certaines collections spécialisées des archives du Web ; au vu de la masse des *ebooks* collectés de cette manière, il ne saurait encore être question de les inclure dans une telle politique. Ce point noir du référencement est un des obstacles techniques les plus frustrants du scénario : capturer une masse aussi importante de données sans être sûr de pouvoir les exploiter à plein, aujourd'hui voire demain. Cependant, une indexation plein texte des contenus — encore limitée actuellement, mais qui est appelée à se développer — relativisera sans doute quelque peu ce problème quand elle atteindra ses pleines possibilités.

Les difficultés de consultation, communes aux deux scénarios, sont également à prendre en compte. La collecte d'*ebooks* implique l'installation de logiciels et une organisation informatique qu'il sera primordial de mettre en place pour permettre l'exploitation des documents collectés.

Conclusion : Le dépôt légal des *ebooks*, une gageure

Nous avons dégagé deux possibilités pour mettre en place le dépôt légal des *ebooks*. Chacune apporte son lot de solutions séduisantes qu'il serait dommage de repousser *a priori* du fait de leurs inconvénients. Une solution hybride, conjuguant dans la mesure du possible le meilleur des deux options, et dont les paramètres pourraient varier en fonction de l'évolution du marché, des technologies et des secteurs éditoriaux, pourrait en définitive être imaginée afin de marier au mieux la solution Extranet et celle du moissonnage du Web.

Sur la couverture de dépôt et l'étendue de la collection patrimoniale : les deux solutions semblent complémentaires sur ce point. La solution moissonnage permet une très large capture des contenus numériques et offre ainsi la possibilité de capturer les livres numériques « dans leur jus » : leurs liens et leurs relations avec tous les autres types de documents en ligne. Du point de vue patrimonial, elle a le mérite de collecter un point capital à côté duquel la solution Extranet risque de passer, à savoir les nouveaux usages qui se développent autour du livre numérique, et leur environnement éditorial. Cet aspect ne doit pas être négligé, car c'est une des missions fondamentales du Dépôt légal. Elle se heurte cependant encore à de nombreuses barrières techniques et juridiques pour les contenus protégés. À l'inverse, la solution Extranet permettrait de collecter les fichiers des *ebooks* payants, mais hors de leur contexte, en s'adressant directement aux éditeurs. L'Extranet pourrait au mieux remplacer peu à peu une mineure partie des collectes courantes par moissonnage, car ces deux options nécessitent un travail similaire d'identification des éditeurs et des titres, et de collaboration avec les déposants.

Sur le référencement et le signalement des ebooks : les possibilités de la solution moissonnage sont limitées, mais non dénuées de perspectives si l'on voulait distinguer les *ebooks* du reste des archives du Web pour proposer un traitement spécifique. La solution Extranet offre en revanche le bénéfice d'un signalement relativement aisé conforme aux normes bibliographiques traditionnelles. Il serait particulièrement avantageux de pouvoir faire des liens entre le catalogue général et les fichiers collectés par l'Extranet comme sur le Web, à l'instar de ce qui se fait déjà entre le catalogue et Gallica. Dans le premier cas, les *ebooks* ont vocation à être signalés parmi les autres ressources du Web, auxquelles ils sont totalement intégrés, dans l'autre, on considère plutôt qu'il faut les insérer dans l'environnement bibliographique de l'édition analogique. Les deux approches ne sont pas incompatibles mais représentent un coût qui ne peut se justifier que par la définition d'un caractère prioritaire pour ce type d'ouvrages en comparaison des autres ressources numériques.

Sur l'aspect de la conservation : destinées toutes les deux à rejoindre le magasin numérique SPAR, ces solutions nécessitent une réflexion quasi-identique sur le traitement et le devenir des fichiers, qui doit de toute façon être menée plus globalement par la BnF dans le contexte de l'essor de ses collections numériques et de son investissement sur le long terme dans la préservation numérique à grande échelle. Les formats, les DRM et les métadonnées sont toutefois trois difficultés à prendre en considération au premier chef ; en dépendent en effet les options de conservation et de consultation. La BnF doit prendre des décisions sur ce point, non seulement pour le bien futur de sa mission bibliothéconomique, mais également pour se positionner comme un interlocuteur fort du secteur de l'édition numérique.

Sur la relation avec les éditeurs : les deux solutions nécessitent, dans des termes un peu différents, un dialogue réel entre la BnF et les éditeurs *d'ebooks*. Ceux-ci semblent peut-être moins directement affectés par la solution moissonnage dans la mesure où elle ne concerne pas aujourd'hui les contenus protégés ; mais elle pourrait nécessiter à terme leur pleine coopération du fait des divers obstacles qui gênent encore les robots moissonneurs et une consultation aisée (DRM notamment). La solution Extranet est sans doute plus favorable à l'ouverture d'un tel dialogue avec les éditeurs, car elle s'inscrirait dans une relation de confiance déjà bien engagée dans le cadre de ce projet et reproduirait le modèle du dépôt (au lieu de la collecte) qui laisse aux éditeurs une meilleure maîtrise des contenus effectivement confiés à la BnF. Elle est par ailleurs la plus à même de valoriser le traitement des *ebooks* par la BnF, grâce aux commodités de gestion et de suivi apportées par des outils tels que le formulaire de dépôt en ligne, mais aussi en termes de publicité pour les éditeurs : les Nouveautés éditeurs ou bien le partenariat avec Gallica constituent des arguments incitatifs au dépôt numérique car ils peuvent offrir une visibilité en ligne aux contenus déposés. Le dialogue avec les éditeurs doit dans tous les cas être soutenu sur les points techniques et ne pourra que bénéficier d'une harmonisation des pratiques et des solutions techniques entre les différents acteurs de la chaîne du livre numérique.

Sur l'aspect valorisation : les deux solutions offrent des possibilités de mise en valeur différentes et complémentaires : d'un côté des parcours ou d'autres produits éditoriaux à inventer au sein des archives du Web, d'un autre la publicité sur Gallica et par le biais des Nouveautés éditeurs, beaucoup plus visible, car en ligne et non cantonnée aux salles de lecture. On pourrait également penser, afin de mettre en exergue ce nouveau support, à une *Bibliographie numérique de la France* (ou à une adjonction de références numériques dans la *Bibliographie nationale de la France*), qui serait un intéressant indicateur de l'état de l'édition numérique nationale. Certes, ce travail repose la question d'une définition stricte et partagée de l'*ebook* (que retenir, qu'exclure de cette bibliographie par principe exhaustive ?) et demanderait un effort de signalement important, mais il s'agirait d'une vitrine et d'un produit d'appel remarquable pour les éditeurs numériques.

Sur l'aspect organisationnel et économique : la solution moissonnage est sans conteste la plus légère et la moins coûteuse, car elle s'inscrit dans un cadre qui arrive déjà à maturité et peut profiter de tous les investissements (techniques et humains) déjà réalisés pour le dépôt légal de l'Internet. Le surcoût à prévoir en termes de stockage et de traitement peut être aisément absorbé par les chaînes de production existantes. La solution Extranet appelle en revanche un développement des ressources humaines dédiées au dépôt légal numérique dans le contexte où la BnF perd chaque année des emplois dans le cadre de la RGPP. Au-delà de la question des effectifs, la création d'une équipe dédiée capable de gérer ce type de dépôt soulève également des enjeux de formation et de qualification. Elle impliquerait sans doute de conjuguer les deux organisations, et de faire évoluer la composition et l'agencement des équipes afin de limiter les coûts humains et de disséminer l'expertise acquise par le service du DLN.

On pourrait enfin envisager une solution intermédiaire ou plutôt mixte, hypothèse transitoire qui serait sans doute l'idéal dans l'état actuel des choses. En effet, il est essentiel d'envisager la mise en place du dépôt légal des *ebooks* d'une manière progressive, qui sache rester souple et évolutive compte tenu des dynamiques encore très incertaines du marché et des technologies employées. Le moissonnage permet de ne pas manquer les « incunables » de l'*ebook*, grâce à son étendue et à la possibilité que le DLN a d'ores et déjà d'intégrer ce flux de production : c'est un acquis précieux. L'Extranet, qui rencontre un vif succès auprès des éditeurs, est quant à lui appelé à se

développer sur le long terme, et le dépôt des livres numériques paraît l'une des évolutions intéressantes de sa trajectoire fonctionnelle.

Il est de toute manière hors de question de passer au tout numérique, et un travail de la BnF en direction de l'édition numérique ne pourrait en tout état de cause aboutir à la substitution totale d'une filière du dépôt légal (papier) par une autre (numérique), quelle qu'elle soit. La coexistence restera longtemps nécessaire, ne serait-ce qu'à cause des contrastes qui existent au sein même du marché de l'édition sur l'adoption du numérique. Les publications de certains domaines de l'édition rejoignent facilement les collections du dépôt légal par papier alors qu'elles auraient peut-être davantage de difficultés ou de réticences à passer à un autre mode de dépôt¹⁸⁵. *A contrario*, d'autres secteurs éditoriaux sont sans doute appelés à se développer particulièrement dans le numérique et à abandonner de plus en plus le papier. L'auto-édition est un bon exemple de cette diversité éditoriale. La coexistence des chaînes du dépôt légal — imprimé et numérique et, à l'intérieur du numérique, des filières moissonnage et Extranet — pourrait ainsi être envisagée comme une combinaison d'approches à moduler non seulement en fonction des évolutions dans le temps, mais aussi en fonction des différents secteurs de l'édition.

On a soulevé plusieurs points dont une réflexion sur le dépôt légal des *ebooks* ne pourra faire l'économie, quel que soit le scénario finalement retenu.

D'abord, la diversité des tendances éditoriales et la nécessaire adaptation de la bibliothèque à l'écosystème complexe de l'édition numérique française impliquent **pour la BnF un travail indispensable de veille, de concertation et de dialogue avec les éditeurs**. Ces échanges doivent intégrer les e-distributeurs, dont l'intervention permettrait sans nul doute un traitement plus complet et plus économique des données, car elle limiterait le nombre des interlocuteurs et favoriserait la normalisation et l'harmonisation des modes et formats d'édition et de diffusion en amont plutôt qu'en aval du dépôt.

Ensuite, il est essentiel de garder à l'esprit que le marché des *ebooks* est encore un univers fluctuant et méconnu ; des orientations hésitantes dictent l'évolution des secteurs éditoriaux, des modèles économiques, des usages ou des volumes, ou le basculement entre l'imprimé et le numérique. Dans ce contexte, **la BnF et les structures qu'elle met en place doivent avant tout rester flexibles et réactives, sans quoi elles encourent le risque de se voir très rapidement périmées**. On ne va vraisemblablement pas vers une disparition massive des supports analogiques, mais plutôt vers une croissance exponentielle du nombre total de publications, ce qui va entraîner pour la bibliothèque un vrai défi d'adaptation. Si ce contexte fait de la flexibilité et de la capacité à évoluer rapidement un premier point de vigilance pour la BnF, il l'interroge aussi sur le caractère exhaustif de sa mission patrimoniale : pourra-t-elle continuer à tout conserver ou devra-t-elle, comme elle l'a déjà engagée avec le dépôt légal de l'Internet, adopter à terme une démarche d'échantillonnage ou de sélection ?

Enfin, la réflexion à conduire en relation avec les éditeurs sur le traitement des *ebooks* permettrait à la BnF de se placer comme un acteur de référence, rôle qu'elle a déjà dans d'autres secteurs où son expertise est reconnue. Dans un domaine encore peu organisé où tout est encore à construire, **la bibliothèque pourrait, à travers la mission patrimoniale de dépôt légal, être reconnue comme un véritable tiers de confiance** et prendre une part plus active à ce titre dans des débats tels que ceux de la gestion des

185. Que l'on pense à Gallimard, déjà mentionné, dont la chaîne de dépôt est mise en place de telle manière que les possibilités de l'Extranet ne l'intéressent pas.

droits ou de l'organisation de l'édition numérique. Ses projets futurs¹⁸⁶ pourront pleinement prendre en compte les *ebooks* comme une nouvelle facette de ses collections numériques, et aborder cet aspect par le dépôt légal semble un bon angle d'attaque.

Pour les bibliothèques, la question du livre numérique soulève plus globalement une réflexion de fond sur la transformation des contenus et des formes du patrimoine. On a vu à quel point l'archivage du Web prenait son importance dans le statut patrimonial de l'environnement éditorial. Internet et ses usages déconstruisent les images fixes que nous mettons derrière les termes de livre, d'éditeur, de librairie, de paysage éditorial. Il en est de même pour les métiers des bibliothèques qui devront s'adapter à ces nouvelles formes du savoir, et développer des compétences qui n'ont rien à envier aux qualifications traditionnelles.

Bibliographie

- ALBANEL (Christine), ROUSSEAU (Aurélien), LADOUARI (Laurent).- *Pour un livre numérique créateur de valeurs : rapport remis au Premier ministre.*- 2009.- 33 p.- Disponible à l'adresse : <http://www.ladocumentationfrancaise.fr/rapports-publics/104000189/>.
- BENHAMOU (Françoise).- *Le livre numérique : ni tout à fait le même, ni tout à fait un autre.*- Lyon : Esprit, mars-avril 2009.- *Esprit*, n° 353.
- BÉCOURT (Daniel).- *Dépôt légal : de l'écrit à l'électronique.*- Paris : Litec, 2001.- 200 p.
- BERMÈS (Emmanuelle).- Blog *Figoblog*.- Disponible à l'adresse : <http://www.figoblog.org/>.- Plusieurs billets sur des aspects techniques plus particulièrement, très clairs.
- BERMÈS (Emmanuelle), DUSSERT CARBONNE (Isabelle), LEDOUX (Thomas), LUPOVICI (Christian).- *La préservation numérique à la Bibliothèque nationale de France : présentation technique et organisationnelle.*- 74^e congrès de l'IFLA, *Bibliothèques sans frontières : naviguer vers une compréhension globale.*- Québec (Canada).- 10-14 août 2008.- 10 p.- Disponible à l'adresse : http://archive.ifla.org/IV/ifla74/papers/084-Bermes_Carbone_Ledoux_Lupovici-trans-fr.pdf.
- BIENVAULT (Hervé).- *Le coût d'un livre numérique : étude réalisée pour le MOTif (observatoire pour le livre et l'écrit en Île-de-France).*- Avril 2010.- 36 p.- Disponible à l'adresse : <http://www.lemotif.fr/fr/etudes-et-analyses/etudes-du-motif/cout-d-un-livre-numerique/>.
- BONENFANT (Bruno).- *Définir une politique d'archivage du web régional en bibliothèque municipale. L'exemple du web forézien.*- Villeurbanne : ENSSIB, 2008.- Mémoire pour le DCB, sous la dir. de Thierry Veyron.
- BORG (Nathalie).- *Régime juridique du livre : de l'imprimé au numérique.*- Septembre 2010.- Mémoire de master 2 Nouvelles technologies de l'information et de la communication, université de Saint-Quentin-en-Yvelines, sous la dir. de Florence Gaullier.- Disponible à l'adresse : <http://www.juriscom.net/uni/visu.php?ID=1255>.
- BOYLAN (Cathy).- *eBooks : acquisition and access.*- 36^e congrès LIBER, *European intergation : conditions and challenges for libraries.*- 3-7 juillet 2007.- 13 p.- Disponible à l'adresse : <http://www.enssib.fr/bibliotheque-numerique/notice-1278>.
- CARBONE (Pierre), CAVALIER (François) (dir.).- *Les collections électroniques, une nouvelle politique documentaire.*- Paris : Electre-Le Cercle de la librairie, 2009.- 311 p.
- CENTRE NATIONAL DU LIVRE, IPSOS MediaCT.- *Publics du livre numérique : enquête menée de septembre 2009 à février 2009.*- Disponible à l'adresse : <http://www.centrenationaldulivre.fr/?Le-livre-sera-t-il-numerique>.
- CHEVRIER (Stéphanie).- *Plaidoyer pour le livre numérique.*- *La vie des idées* [en ligne], 6 octobre 2008.- Disponible à l'adresse : <http://www.laviedesidees.fr/Plaidoyer-pour-le-livre-numerique.html> [consulté le 6 décembre 2010].
- CLAYSSSEN (Virginie).- Blog « teXtes ». Disponible à l'adresse : <http://www.archicampus.net/wordpress/>.- Principalement centré sur les questions du numérique, notamment en lien avec le livre.
- COHEN-HADRIA (Martine).- *Étonnante collection Charles Cros.*- Paris : BnF, octobre-décembre 2006.- *Chroniques de la Bibliothèque nationale de France*, n° 37.- Disponible

à l'adresse :

<http://chroniques.bnf.fr/archives/decembre2006/numero_courant/coulisse/charles_cros.htm> (consulté le 16 décembre 2010).

CONFÉRENCE ANNUELLE DES BIBLIOTHÉCAIRES NATIONAUX EUROPÉENS.-
Déclaration sur le développement et la création des mécanismes du dépôt volontaire des publications électroniques.- Luxembourg, septembre 2005.- 7 p.

DAVAL (Mathias), DOUINE (Rémi).- *Ebookz, l'offre légale et illégale de livres numériques.*- Paris : MOTif, mai-juin 2010.- 33 p.- Disponible à l'adresse : http://www.lemotif.fr/fichier/motif_fichier/171/fichier_fichier_.l.offre.de.livres.numa.riques.en.france.le.motif.pdf (consulté le 2 janvier 2010).

DELANGLE (Olivier).- *Le département du dépôt légal à la Bibliothèque nationale de France : pour une exhaustivité raisonnée.*- Villeurbanne : ENSSIB, 2003.- 61-XII p.- Mémoire pour le DCB sous la dir. de Sophie Mazens.

DEVIDAL (Maryline).- *Stratégies pour un dépôt légal des périodiques en ligne – Dossier d'aide à la décision.*- Villeurbanne : ENSSIB, 2010.- FIBE, Projet professionnel personnel sous la dir. d'Alix Bruys.

ÉDUCNET.- *Livre électronique, livre numérique (e-book) : dossier.*- Juin 2010.- Disponible à l'adresse : <http://www.educnet.education.fr/dossier/livrelec>.

ÉPRON (Benoît) (coord.).- Dossier : Les livres numérique : de nouveaux enjeux professionnels.- Paris : Association française des documentalistes et des bibliothécaires spécialisés, mai 2010.- *Documentaliste - Sciences de l'information*, vol. 47, n° 2.- p. 26-69.

DUCHARME (Christian) (resp. sc.).- *Les stratégies éditoriales à l'heure du numérique et leurs conséquences sur l'offre des bibliothèques* : journée d'étude.- ENSSIB, 18-19 juin 2003.- 17 p.- Disponible à l'adresse : http://partages.univ-rennes1.fr/files/partages/Services/SCD/SiteWeb/livre_elec/Seminaire_Strategies_editoriales_a_lheure_du_numerique_Lyon_2003.pdf.

GAME (Valérie).- *Le dépôt légal : dernières évolutions législatives et réglementaires* : journées autour du Réseau des bibliothèques de dépôt légal imprimeur.- BnF, 22-23 janvier 2009.- 6 p.- Disponible à l'adresse : http://www.bnf.fr/documents/dl_loi.pdf.

GAYMARD (Hervé).- *Situation du livre : évaluation de la loi relative au prix du livre et questions prospectives* : rapport au MCC.- Mars 2009.- 122 p.- Disponible à l'adresse : <http://www.centrenationaldulivre.fr/?Situation-du-livre-Evaluation-de>.

GUILLAUD (Hubert).- Blog *La feuille : l'édition à l'heure de l'innovation.*- Disponible à l'adresse : <http://lafeuille.blog.lemonde.fr/>.- Plusieurs articles intéressant l'édition numérique et ses modèles.

ILLIEN (Gildas).- *Le dépôt légal de l'internet en pratique : les moissonneurs du Web.*- Paris : BBF, 2008.- *BBF*, vol. 53, n° 6, p. 20-27.- Disponible à l'adresse : <http://bbf.enssib.fr/consulter/bbf-2008-06-0020-004> (consulté le 7 décembre 2010).

ILLIEN (Gildas).- *L'archivage d'internet, un défi pour les décideurs et les bibliothécaires : scénarios d'organisation et d'évaluation – L'expérience du consortium IPC et de la BnF.*- 74^e congrès de l'IFLA, *Bibliothèques sans frontières : naviguer vers une compréhension globale.*- Québec (Canada), 10-14 août 2008.- 14 p.- Disponible à l'adresse : <http://archive.ifla.org/IV/ifla74/papers/107-Illien-fr.pdf>.

ILLIEN (Gildas).- *Qualité et massification du patrimoine numérique : le cas du dépôt légal du web.*- Journée AFNOR-BnF *Services documentaires : quelle qualité pour quels*

- clients ?*, 7 juin 2010.- 36 p.- Support disponible à l'adresse : http://www.bnf.fr/documents/qualite_dl_web.pdf.
- KIM (Yunhyong).- L'extraction automatisée des métadonnées sémantiques.- *Briefing Papers*, 22 octobre 2008.- Trad. Jean-Pierre Hazel Teil.- 2 p.- Disponible à l'adresse : http://www.digitalpreservationeurope.eu/publications/briefs/11_FRENCH.pdf (consulté le 27 décembre 2010).
- KONRATH (Joseph Andrew).- *ebooks and the ease of self-publishing*.- *The Huffington Post* [en ligne], 16 octobre 2010.- Disponible à l'adresse : http://www.huffingtonpost.com/ja-konrath/ebooks-and-self-publishing_b_764516.html (consulté le 27 novembre 2010).
- KREBS (Constance).- Du livre électronique à l'encre électronique: nouveau papier, nouveau livre ?.- Paris : BBF, 2006.- *BBF*, vol. 51, n° 4, p. 38-41.- Disponible à l'adresse : <http://bbf.enssib.fr/consulter/bbf-2006-04-0038-006> (consulté le 27 novembre 2010)
- LEBERT (Marie).- *Une courte histoire de l'ebook*.- NEF, Université de Toronto, 2009.- Disponible à l'adresse : <http://www.etudes-francaises.net/dossiers/ebook.htm>.
- LEMONDE.FR, AFP, Reuters.- De nouvelles liseuses électroniques arrivent sur le marché français.- Paris : Société éditrice du Monde, 22 octobre 2010.- *Le Monde.fr* [en ligne].- Disponible à l'adresse : http://www.lemonde.fr/technologies/article/2010/10/22/de-nouvelles-liseuses-electroniques-arrivent-sur-le-marche-francais_1430059_651865.html (consulté le 19 novembre 2010).
- MAUREL (Lionel).- *Panorama des systèmes de métadonnées juridiques et de leurs applications en bibliothèque numérique*.- 2006.- 20 p.- Disponible à l'adresse : http://archivesic.ccsd.cnrs.fr/sic_00130222/fr/ (consulté le 20 décembre 2010).
- PATINO (Bruno).- *Le devenir numérique de l'édition : du livre objet au livre droit : rapport au MCC*.- Paris : La Documentation française, 2008.- 88 p.- Série des rapports officiels.- Disponible à l'adresse : <http://lesrapports.ladocumentationfrancaise.fr/BRP/084000381/0000.pdf>.
- PILLORGET (Stéphane).- Les e-books à la BnF – Le projet Gallica.- Journée GO !Doc E-book : nouveaux supports, nouveaux usages, 15 septembre 2009.- 29 p.- Support d'intervention disponible à l'adresse : www.godoc.u-psud.fr/IMG/pdf/Pillorget.pdf.
- PROST (Bernard).- *Rapport d'étude sur l'édition numérique de livres scientifiques et techniques, l'éditeur des années 2010* : étude réalisée par QUAE avec le soutien du ministère de l'Enseignement supérieur et de la Recherche.- 2008.- 95 p.- Disponible à l'adresse : http://media.education.gouv.fr/file/2008/41/8/rapporteditionnumerique_28418.pdf.
- Table ronde *Stratégies éditoriales à l'heure du livre électronique*.- Médiathèque de l'Espace Landowski, Boulogne Billancourt, 9 février 2009.- Compte rendu disponible sur le blog du *BBF* à l'adresse : <http://bbf.enssib.fr/blog/2009/02/25/le-livre-numerique-table-ronde-du-9-fevrier-2009-a-boulogne-billancourt> (consulté le 28 novembre 2010).
- VASSILIOU (Magda), ROWLEY (Jennifer).- Progressing the definition of « e-book ».- *Library Hi-Tech* [en ligne], 2008, vol. 26, n° 3, p. 360.- Disponible à l'adresse : <http://www.emeraldinsight.com/journals.htm?issn=0737-8831&volume=26&issue=3&PHPSESSID=0shpla9aqbrikqqr377q9ad6p2>.

Textes de loi

Définition du livre électronique de la commission générale de terminologie et de néologie, publiée au JO du 18 janvier 2005.

Loi no 2006-961 du 1^{er} août 2006 sur le droit d'auteur et les droits voisins dans la société de l'information (DADVSI). Disponible à l'adresse : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000266350>

Code du patrimoine, titre III sur le dépôt légal, art. L. art. 131-1 à 133-1. Disponible à l'adresse : <http://www.legifrance.gouv.fr/affichCode.do?cidTexte=LEGITEXT000006074236>

Décret n° 2006-696 du 13 juin 2006 modifiant le décret n° 93-1429 du 31 décembre 1993 relatif au dépôt légal. Disponible à l'adresse : <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000816574&dateTexte=>

Entretiens

Entretiens avec Gildas Illien, les 18 juin, 31 août (courriel), 14 octobre, 5 novembre, 3 décembre 2010 et 19 janvier 2011 (téléphone).

Entretien téléphonique avec Emmanuelle Bermès le 27 août 2010.

Entretien avec Anne Bommier-Chasles le 15 novembre 2010.

Entretien (courriel) avec Elodie Bertrand le 13 décembre 2010.

Table des annexes

ANNEXE 1 : LES PRINCIPAUX FORMATS DE LIVRES NUMERIQUES ACTUELLEMENT DISPONIBLES	73
ANNEXE 2 : ÉVALUATION PAR RAYONS	74
ANNEXE 3 : FORMULAIRES DE DEPOT.....	78
ANNEXE 4 : LES FORMATS DE METADONNEES.....	82
ANNEXE 5 : PROPOSITION DE SITES CIBLES POUR UNE COLLECTE COURANTE	84

Annexe 1 : Les principaux formats de livres numériques actuellement disponibles¹⁸⁶

	Adobe PDF	ePub	Mobipocket PRC	Microsoft LIT	MP3	Microsoft WMA
Extension	PDF	ePub	PRC	LIT	MP3	WMA
Quel logiciel ?	Adobe Digital Editions	Adobe Digital Editions	Mobipocket Reader	Microsoft Reader	Windows Media Player, itunes, Winamp, Real One Player	Windows Media Player
PC	Oui	Oui	Oui	Oui	Oui	Oui
Mac	Oui pour Mac OS X v10.4.10 ou v10.5	Oui pour Mac OS X v10.4.10 ou v10.5	Non mais vous pouvez les télécharger sur votre PDA via votre Mac	Non	Oui	Non
PDA	Non	Non	Oui, la plupart des PDA sont compatibles	Oui si votre PDA est un Pocket PC 2002 / 2003	Oui	Oui voir les appareils compatibles
Linux	Non	Non	Non	Non	Oui	Non
eReaders	Sony Reader PRS-505	Sony Reader PRS-505	Cybook, iLiad, Pepper-Pad	Non	Oui	Non
Imprimable	Selon autorisation	Non	Non	Non	Non	Non
Combien de titres sont disponibles sur Numilog ?	28652	12276		209		790

186. Ce tableau est reproduit dans son état au 20 novembre 2010, et est consultable sur le site de Numilog : <<http://www.numilog.fr/aide.asp>>.

Annexe 2 : Évaluation par rayons

Remarque : ces chiffres¹⁸⁷ sont à valeur indicative, car des ouvrages reviennent d'une catégorie à l'autre et sont donc pris en compte deux fois. Les catégories les plus importantes sont indiquées en gras. Nous n'avons retenu que les e-distributeurs Numilog, Eyrolles et Mobipocket car ce sont les seuls à afficher le nombre d'ouvrages contenus dans les « rayons » (contrairement à Fnac.com par exemple), et qu'ils restent assez généralistes.

NUMILOG		
Savoirs	Concours-Méthodes	217
	Droit	398
	Entreprise	1251
	Économie	683
	Essais	1465
	Histoire-Géo	1203
	Informatique	594
	Langues	33
	Lettres	300+ (27 p.) ¹⁸⁸
	Pédagogie	19
	Sciences humaines	1375
	Sciences et techniques	1416
Littérature	Religions et spiritualités	673
	BD	33
	Classiques	300+ (187 p.)
	Érotisme	754
	Fantastique et SF	258
	Humour	33
	Jeunesse	300+ (31 p.)
	Littérature sentimentale	221
	Mythologies et contes	8
	Mémoires et biographies	300+ (65 p.)
	Policiers et thrillers	300+ (47 p.)
	Poésie	240
	Romans d'aventure	18
	Romans et nouvelles	3080
	Théâtre	216
Pratique	Apprentissage des langues	47
	Bien-être et développement personnel	300+ (38 p.)
	Guides pratiques et vie quotidienne	269

187. Chiffres au 20 décembre 2010.

188. Un bug dans l'affichage du catalogue de ce site ne permet de voir que les 21 premières pages de chaque catégorie (et donc les 300 premiers ouvrages). Quand les chiffres étaient indiqués, on en a déduit le nombre d'ouvrages de la catégorie, mais quand ce n'était pas le cas, on a fait une estimation selon le nombre de pages totales indiqué.

	Loisirs	79
	Métiers et formations	274
	Sports	40
	Voyages	92
Arts	Architecture	12
	Arts, ouvrages généraux	36
	Beaux-Arts, peinture	36
	Cinéma	37
	Musique	87
	Photographie	40
	Urbanisme et art du paysage	11
	Dictionnaires	44
	Livres offerts	143
	Audio	300+ (96 p.)
EYROLLES - IZIBOOK		
Littérature		41
Informatique	Développement d'applications	75
	Développement informatique	68
	Informatique	144
	Internet	66
	Langages informatiques	49
	Programmation	64
	Techniques	143
	Web	66
	Total	159
Graphisme et Web	Audiovisuel	6
	Design	11
	Graphisme	11
	Images	11
	Informatique	9
	Logiciel	7
	Multimédia	7
	Techniques	10
	Total	16
Techniques photo	Audiovisuel	15
	Images	16
	Informatique	13
	Multimédia	13
	Numérique	8
	Photographie	16
	Photographie numérique	8
	Techniques	16
	Total	20
Entreprise	Communication	70

	Efficacité professionnelle	127
	Entreprise	329
	Formation	67
	Management	194
	Sciences humaines	108
	Sciences économiques	96
	Économie d'entreprise	74
	Total	406
Sciences humaines	Entreprise	69
	Europe	42
	France	42
	Histoire	43
	Management	43
	Sciences humaines	134
	Sciences économiques	35
	Vie pratique	53
	Total	150
Vie pratique	Bien-être	60
	BTP et Construction	42
	Développement personnel	46
	Immobilier	44
	Loisirs et vie quotidienne	86
	Sciences humaines	56
	Techniques	49
	Vie pratique	173
	Total	175
BTP et Construction	Bricolage	18
	BTP et Construction	34
	Construction	34
	Immobilier	34
	Maison	17
	Maison individuelle	18
	Techniques	34
	Vie pratique	19
	Total	53
Artisanat et décoration	Aménagement	1
	Bijoux	1
	Cheminée	1
	Dessin	1
	Jardin	1
	Mosaïque	1
	Peinture	1
	Vannerie	2
	Total	9
Gratuits		12

MOBIPOCKET		
Romans	Général	66
	SF	150
	Romans d'amour	18
	Fantastique	33
	Suspense et Thrillers	44
	Aventure	102
	Policier et Crime	49
	Érotisme	157
	Littérature classique	76
	Romans historiques	36
	Jeunesse	50
	Romans	1648
	Épouvante	5
	Humour	10
	BD	24
	Fiction religieuse	1
	Contes et fables	78
	Nouvelles	69
	Poésie	313
	San Antonio	48
Théâtre	103	
Références	Médical	23
	Dictionnaires	47
	Religion	28
	Développement personnel	14
	Psychologie	12
	Santé et médecine	3
	Histoire	186
	Politique	30
	Biographies	36
	Actualité	16
	Voyages et Tourisme	24
	Famille	3
	Cuisine	5
	Informatique	8
	Sciences	38
	Livres de référence	9
	Arts	9
	Philosophie	238
	Sports	2
	Linguistique	5
Finances personnelles	1	

Annexe 3 : Formulaire de dépôt

FORMULAIRE DE DEPOT POUR LES IMPRIMES



Bibliothèque nationale de France

déclaration de dépôt légal - éditeur livres
et autres documents non périodiques

cerfa n° 10069*03 | Code du patrimoine art. L 131-1 à L 133-1 - Décret n° 93-1429 du 31 décembre 1993 (arrêtés du 12 janvier 1995)

réserve à l'administration

A Identification du déposant

téléphone _____

courriel _____

signature du déposant _____

à _____

le _____

1 Nom (ou raison sociale) et adresse*

* La loi N° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux données nominatives portées sur ce formulaire. Elle garantit un droit d'accès et de modification pour ces données auprès de la Bibliothèque nationale de France. Si vous ne souhaitez pas que votre adresse personnelle apparaisse dans les produits bibliographiques diffusés, en particulier sur le site Internet de la BnF, veuillez cocher la case ci-contre

B Descriptif du document déposé

2 ISBN _____

3 auteur[s] nom, prénoms _____ date de naissance _____ pseudonyme _____

4 document titre _____

titre de la collection _____

numéro dans la collection _____

5 édition nouveauté nouvelle édition broché relié

format en cm _____

nbre de pages _____

7 chiffre déclaré du tirage _____

prix de vente au public _____

dépôt obligatoire en 2 exemplaires (1 exemplaire si le tirage est inférieur à 300 exemplaires)

6 nom (ou raison sociale) et adresse de l'imprimeur _____

nom (ou raison sociale) et adresse du dernier façonnier _____

nombre d'exemplaires déposés 1 2

8 date de mise à disposition du public _____

C Lieu de dépôt

9 Bibliothèque nationale de France
Dépôt légal - Livres
Quai François Mauriac
75706 Paris Cedex 13

Téléphone : 01 53 79 43 37
Télécopie : 01 53 79 45 60
depot.legal.livres@bnf.fr

Déclaration à compléter et à joindre en 3 exemplaires à chaque titre déposé **T. S. V. P.**

FORMULAIRE DE DEPOT POUR LES DOCUMENTS INFORMATIQUES (LOGICIELS ET BASES DE DONNEES)

{BnF} déclaration de dépôt légal - logiciels et bases de données

cerfa
n° 10063*04

Code du patrimoine art. L. 131-1 à L. 133-1 - Décret n° 93-1429 du 31 décembre 1993
(arrêté du 12 janvier 1995)

réserve à l'administration

A Identification du déposant

téléphone _____
courriel _____
nom du correspondant _____
signature du déposant _____
à _____
le _____

1 — Nom (ou raison sociale) et adresse*

* La loi N° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés s'applique aux données nominatives portées sur ce formulaire. Elle garantit un droit d'accès et de rectification pour ces données auprès de la Bibliothèque nationale de France. Si vous ne souhaitez pas que votre adresse personnelle apparaisse dans les produits bibliographiques diffusés, en particulier sur le site Internet de la BnF, veuillez cocher la case ci-contre

2 — fonction éditeur producteur commanditaire distributeur

B Descriptif du document déposé dépôt obligatoire en 2 exemplaires

titre du document _____

3 — auteur[s] nom, prénoms

4 — nom (ou raison sociale) et adresse des responsables de la publication

5 — titre de la collection n° dans la collection

6 — édition nouveauté nouvelle édition 7 — nb d'exemplaires déposés 1 2

8 — chiffre déclaré du tirage ou de l'importation

prix de vente au public

date de mise à disposition du public

9 — nb d'unités physiques

environnement matériel et logiciel requis

nature du support

indication de la documentation fournie

10 — système de protection

C Lieu de dépôt

11 — Bibliothèque nationale de France
Dépôt légal - Multimédias T3N3
Quai François Mauriac - 75706 Paris Cedex 13

Téléphone : 01 53 79 53 18
Télécopie : 01 53 79 47 21
depotlegal.multimedia@bnf.fr

Déclaration à compléter et à joindre à chaque document déposé T.S.V.P.

Annexe 4 : Les formats de métadonnées

LES ELEMENTS DU FORMAT DUBLIN CORE

1. Titre : Titre principal du document.
2. Créateur : Nom de la personne, de l'organisation ou du service à l'origine de la rédaction du document.
3. Sujet(s) ou mots clés : Mots-clefs, phrases de résumé, ou codes de classement.
4. Description : Résumé, table des matières, ou texte libre.
5. Éditeur : Nom de la personne, de l'organisation ou du service à l'origine de la publication du document.
6. Contributeur : Nom d'une personne, d'une organisation ou d'un service qui contribue ou a contribué à l'élaboration du document. Chaque contributeur fait l'objet d'un élément séparé.
7. Date : Date d'un évènement dans le cycle de vie du document.
8. Type de ressource : Genre du contenu.
9. Format : Type MIME, ou format physique du document.
10. Identifiant de la ressource : Identificateur non ambigu : il est recommandé d'utiliser un système de référencement précis, afin que l'identifiant soit unique au sein du site, par exemple les URI¹⁸⁹ ou les numéros ISBN. Raffinement : *Is Available At* (Est disponible à/sur).
11. Source : Ressource dont dérive le document : le document peut découler en totalité ou en partie de la ressource en question. Il est recommandé d'utiliser une dénomination formelle des ressources, par exemple leur URI.
12. Langue.
13. Relation : Lien avec d'autres ressources. De nombreux raffinements permettent d'établir des liens précis, par exemple de version, de chapitres, de standard, etc.
14. Couverture : couverture spatiale (point géographique, pays, régions, noms de lieux) ou temporelle.
15. Gestion des droits : Droits de propriété intellectuelle, Copyright, droits de propriété divers.

Tous les éléments ne sont d'ailleurs pas obligatoires et peuvent être répétés, et leur emploi est défini en fonction de l'utilisation que l'on souhaite faire du fichier concerné.

L'utilisation des « raffinements », c'est-à-dire la précision de la signification de certains éléments, détermine l'appellation de Dublin Core qualifié. Cette utilisation est facultative. Il faut préciser que le protocole OAI-PMH¹⁹⁰ repose sur du Dublin Core non qualifié. Ce dernier est devenu en 2003 une norme internationale (ISO 15836). Le Dublin Core non qualifié est utilisé¹⁹¹ à la BnF pour gérer et cataloguer les signets, et accroître sur Internet la visibilité des catalogues et des collections numérisées, de Gallica notamment. Il s'agit d'aider les acteurs qui utilisent d'autres formats que les formats bibliographiques des bibliothèques à référencer les données de la BnF.

189. *Uniform Resource Identifier*.

190. *Open Archive Initiative Protocol for Metadata Harvesting* : protocole pour la collecte de métadonnées de l'initiative pour les archives ouvertes.

191. Voir le guide d'utilisation du Dublin Core à la BnF, disponible à l'adresse : http://bibnum.bnf.fr/oai/GuideOAI_2008_09_08.pdf

LES ELEMENTS DU FORMAT ONIX

La liste des éléments renseignés dans le format ONIX est bien plus longue que pour le format DC. Nous ne la développerons pas ici en détail. Sur le site d'EDITEUR¹⁹², responsable du développement de ONIX Livres, basé à Londres, on trouvera une traduction de la version 3.0 réalisée sous le contrôle du groupe ONIX France et avec le soutien du Cercle de la Librairie à l'adresse : <http://www.editeur.org/93/Release-3.0-Downloads/>.

Les éléments sont groupés et ordonnés dans des blocs, suivant des paramètres d'identification obligatoires :

Bloc 1 : description bibliographique,

Bloc 2 : informations utiles à la commercialisation,

Bloc 3 : détail du contenu : données structurées sur les chapitres ou d'autres éléments de contenu,

Bloc 4 : conditions de publication (éditeur, etc.),

Bloc 5 : matériels liés (œuvres et autres produits en rapport avec le produit décrit),

Bloc 6 : conditions d'approvisionnement, sur des marchés spécifiés le cas échéant.

192. <<http://www.editeur.org>>

Annexe 5 : Proposition de sites cibles pour une collecte courante

Nom du producteur	Adresse à cibler	Remarques
AveComics	http://www.avecomics.com/#/	Spécialisé dans la BD numérique, sur supports multiples. Une trentaine d'éditeurs ¹⁹³ .
The Book edition*	http://www.thebookedition.com/	Auto-édition.
Cyberlibris	http://www.cyberlibris.com/	Avec ses sous-domaines (sciences, scholarvox, etc.).
Dalloz	http://www.dalloz-achatenligne.fr/dalloz/	Spécialisé en droit.
Dialogues	http://www.librairiedialogues.fr/	Fournisseurs : Eden Livres, Editis, Immatériel, Numilog, Entrepôt numérique. Mais aussi éditeur.
Didactibook	http://www.didactibook.com/	Livres pratiques. 15 éditeurs.
Digibidi	http://www.digibidi.com/	Bandes dessinées. Une trentaine d'éditeurs.
La Documentation française	http://www.ladocumentationfrancaise.fr/acheter/	
Ebooks gratuits*	http://www.ebooks-gratuit.com/	Uniquement du libre de droit.
Ebookslibresetgratuits.com	http://www.ebooksgratuits.com/	
Ebook Livre	http://www.ebook-livre.com/	Auto-édition.
Eden Livres	http://www.edenlivres.fr/	Groupes Flammarion, Gallimard et La Martinière. Environ 4.000 références, issues du fonds de ces 3 groupes + des nouveautés. Pas de livres à contenu enrichi.
Éditeur indépendant	http://www.editeurindependant.com/doc	Spécialiste de la publication de premiers romans ¹⁹⁴ .
Eyrolles	http://izibook.eyrolles.com/	Eyrolles + 10 éditeurs. Particulièrement intéressant : les « purs ebooks » : http://izibook.eyrolles.com/purs_ebooks .

193. Ces chiffres sont issus de l'étude du MOTif, *Ebookz...*, op.cit., p. 16-17.

194. Sélection mensuelle d'un comité de lecture pour un accès à une collaboration sous contrat d'édition. Dépôt légal organisé : <<http://www.editeurindependant.com/publier>>. Papier et numérique.

Fnac.com	http://www.fnac.com/telecharger-ebook.asp	Les offres papier (VPC) et numérique ne sont pas séparées.
Gallimard ¹⁹⁵	http://www.gallimard.fr/ http://www.librairie-gallimard.com/livrel.php?ALIS=10f997b6379ca60a78237d0e0a2dc21d	Vente du fonds Gallimard disponible en version numérique.
Éditions Harlequin	http://numerique.harlequin.fr/	Offre comprise dans Numilog, à voir donc.
I-Kiosque	http://www.i-kiosque.fr/doc	46 éditeurs.
Immatériel	http://librairie.immateriel.fr/	25 éditeurs.
Izibook PUG ¹⁹⁶	http://www.pug.fr/ Inédits numériques : http://www.pug.fr/purs_ebooks	Catalogues papier et numérique mêlés.
Izneo	http://www.izneo.com/	Bande dessinée, à l'achat ou à la location. 12 éditeurs.
L'Harmattan	http://www.editions-harmattan.fr/index.asp?navig=catalogue&obj=result&ntable=0&ebook=1&orderby=date&ordermode=DESC	<i>Ebooks</i> de L'Harmattan.
Livres pour tous	http://www.livrespourtous.com/	
Lulu*	http://www.lulu.com/fr/	Auto-édition
Mobipocket*	http://www.mobipocket.com/en/ebooks/default.asp?Language=FR/	
Numilog*	http://www.numilog.fr/	Groupe Hachette ; 150 éditeurs.
Pitbook	http://www.pitbook.com/	
Publie.net	http://www.publie.net/	
Zebook	http://www.zebook-store.com/	

195. Bien que ne mettant pas directement en vente des livres numériques, ce site est particulièrement intéressant dans le sens où il propose des mini-sites autour de certains ouvrages. Un lien est fait à partir de ces mini-sites soit vers une commande papier, soit vers des e-distributeurs (Amazon, Fnac, Place des libraires).

196. Presses universitaires de Grenoble.