



HAL
open science

Le dépôt légal des livres numériques en France et ses conséquences sur le catalogage et la Bibliographie nationale

Sophie Derrot, Mathilde Koskas

► **To cite this version:**

Sophie Derrot, Mathilde Koskas. Le dépôt légal des livres numériques en France et ses conséquences sur le catalogage et la Bibliographie nationale. IFLA World Library and Information Congress, IFLA, Aug 2015, Cape Town, Afrique du Sud. hal-01248890

HAL Id: hal-01248890

<https://bnf.hal.science/hal-01248890>

Submitted on 30 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le dépôt légal des livres numériques en France et ses conséquences sur le catalogage et la Bibliographie nationale

French translation of the original paper / Traduction française de l'article original : "Every Reader a Library, Every Library its Reader: Designing Responsive Libraries for our Communities".

Traduit par: Ariane Bouchard, Bibliothèque nationale de France, Paris

Sophie Derrot

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

Mathilde Koskas

Département du dépôt légal, Bibliothèque nationale de France, Paris (conférencière)
mathilde.koskas@bnf.fr



This is a french translation of "Legal deposit of ebooks in France and its bearings on cataloguing and the National Bibliography" copyright © 2015 by Ariane Bouchard. This work is made available under the terms of the Creative Commons Attribution

3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Résumé :

La loi française sur le dépôt légal couvre les sites et les contenus en ligne tout comme les livres numériques. Elle n'impose pas l'obligation de produire une bibliographie dans sa forme traditionnelle, l'indexation étant suffisante. Mais malgré leur caractère innovant, les livres numériques restent des livres et leurs métadonnées bibliographiques sont plus proches de celles du livre imprimé que de celles qui permettent l'indexation des archives du web. Pour monter un circuit complet de dépôt du livre numérique, la BnF bénéficie à la fois de son expérience dans le domaine des documents numériques et de la tradition du dépôt légal. Cet article a pour but de présenter les questions auxquelles la BnF est confrontée en matière de catalogage des livres numériques et de gestion de leurs métadonnées, et les solutions qui émergent.

Mots-clés : livres numériques, dépôt légal, bibliographie nationale, normes de catalogage

La nature du livre numérique est double, entre le document numérique et le livre traditionnel, et son statut est ambigu, comme le montre le débat sur sa taxation au sein de l'Union européenne. La Bibliothèque nationale de France a pour mission de collecter les contenus en ligne tout comme les livres imprimés. Pendant longtemps, ces contenus en ligne ont surtout été les sites web, mais depuis 2012, la BnF travaille sur un dépôt légal propre aux livres

numériques. Ce type de dépôt légal tombe sous le coup de la loi de 2006 sur le numérique, mais le circuit qui permettra à la BnF de mieux traiter les documents et de faciliter le dépôt des éditeurs est très proche de celui du dépôt légal traditionnel. Le principal problème est que la loi sur le dépôt légal numérique diffère en plusieurs points de celle sur le dépôt légal imprimé : il n'y a ainsi aucune obligation d'exhaustivité ni de production d'une bibliographie nationale. Ces différences doivent être prises en compte car elles tracent une frontière entre le traitement des livres numériques et celui des livres imprimés.

La dématérialisation du dépôt légal a débuté lorsque les éditeurs ont eu la possibilité de remplir leurs déclarations en ligne, grâce à un Extranet spécifique. Plus récemment, la possibilité leur a été offerte de mandater leur distributeur pour l'envoi d'un flux de métadonnées en guise de déclaration d'un livre imprimé. La prochaine étape est l'automatisation du traitement des livres numériques et de leurs métadonnées versés à la BnF. Pour réussir cette évolution du dépôt légal, la BnF doit prendre position sur des questions bibliographiques importantes et résoudre les contradictions entre la théorie bibliographique et l'organisation d'une telle institution. De bien des façons, le postulat de la proximité avec le livre imprimé a ses limites, que nous devons prendre en compte lorsque nous décidons de la meilleure façon de créer et disséminer les notices bibliographiques des livres numériques.

Travailler avec les métadonnées des éditeurs

Le projet d'un dépôt légal des livres numériques a débuté en 2012 par un dialogue avec notre partenaire du monde de l'édition, le Syndicat national de l'édition. La première phase a été importante car bien des aspects du marché du livre numérique étaient assez mal connus de la BnF : combien de livres numériques sont publiés par an ? Quels sont les formats les plus communs ? Comment fonctionne le circuit de l'édition et de la commercialisation ? Toutes ces questions sont cruciales pour l'organisation d'un dépôt légal viable. L'une des premières questions à laquelle nous avons dû faire face durant ces entretiens concernait les métadonnées : comment l'information à propos d'un livre numérique va-t-elle de l'auteur et de l'éditeur au vendeur et au consommateur ? À mesure que notre connaissance du circuit augmentait, nous nous sommes rendu compte que certains acteurs majeurs ajoutaient des métadonnées au cours du traitement et que le besoin d'avoir une norme de transfert des métadonnées préexistait à la demande de la BnF.

La norme la plus utilisée dans le monde de l'édition (en France et dans le monde) est l'une des normes ONIX dédiée aux livres. *ONIX for books* a été développée à l'origine par EDItEUR et est gérée aujourd'hui par cette entreprise et un comité de pilotage international¹. Elle est conçue pour assurer la communication entre ordinateurs en langage XML. Un fichier ONIX permet de décrire les livres imprimés comme les livres numériques et peut exprimer des données bibliographiques, techniques et commerciales grâce à une grande variété de champs décrits de façon précise dans une documentation disponible en ligne et régulièrement mise à jour. Au sein du marché français du livre numérique, le fichier ONIX décrivant un livre est créé par le distributeur en fonction des informations fournies par l'éditeur.

La BnF sait déjà traiter les fichiers ONIX des livres imprimés. Depuis fin 2014, les éditeurs ont la possibilité de faire leur déclaration de dépôt légal directement par un flux ONIX fourni par leurs distributeurs à la BnF. Les livres sont envoyés séparément et, quand ils arrivent à la BnF, les déclarations en ligne sont vérifiées automatiquement puis transférées à l'application

¹ <http://www.editeur.org/8/ONIX/>

de catalogage afin d'être traitées par les catalogueurs. Cette expérience a été très utile pour la mise en place du circuit du dépôt légal des livres numériques.

Tous nos partenaires utilisent des fichiers ONIX 3.0 afin de transférer des métadonnées de livres numériques du distributeur au vendeur en ligne. Ils nous ont transmis des échantillons de leur production, afin que nous puissions avoir une idée de la qualité de leur fichier ONIX. Pendant ce temps, nous avons mené une réflexion sur la forme que devait prendre une notice de livre numérique dans le catalogue général de la BnF². Nous avons comparé les informations des deux sources et ainsi créé un modèle ONIX idéal, avec des champs obligatoires, facultatifs ou souhaités. Ce modèle a été envoyé à nos partenaires ; il tend à évoluer avec la réflexion en cours sur le sujet et les informations fournies par les distributeurs.

Échantillon du modèle ONIX de la BnF, concernant le format du livre numérique et son titre

Champs ONIX 3	Valeurs et descriptions ³	Standard ONIX ⁴	Demande BnF
<DescriptiveDetail>		O	O
<ProductComposition>	00 : produit composé d'un seul article	O	O
	10 : produit composé de plusieurs articles		
	11 : collection composée de plusieurs articles, fournis séparément		
<ProductForm>	EA : contenu numérique (fourni par voie électronique)	O	O
	EB : contenu numérique (accessible par téléchargement et en ligne)		
	EC : contenu numérique uniquement accessible en ligne		
	ED : contenu numérique uniquement téléchargeable		
<ProductFormDetail>	E101 : EPUB		O
	E107 : PDF		
	E108 : PDF/A		
	E200 (si E101) : redimensionnable		
	E201 (si E101) : format fixe		

[...]

<TitleDetail>		R	R
<TitleType>	01 : titre significatif	O	O
	03 : titre original		
	Les autres valeurs sont facultatives.		
<TitleElement>			
<TitleElementLevel>	01 : produit	O	O
<TitleText>	Libellé de titre	O	O
<TitlePrefix>	Préfixe	F	O si préfixe
<TitleWithoutPrefix>	Titre sans le préfixe	F	O si préfixe
<Subtitle>			O si préfixe
</TitleElement>			
</TitleDetail>			

². La BnF achète déjà des livres numériques (et des accès). Cependant, ils ne sont pas signalés dans le Catalogue général mais dans une application propre aux ressources numériques telles que les bases de données ou les périodiques en ligne.

³. Ces valeurs renvoient à la liste des codes attachée à la norme.

⁴. O pour Obligatoire ; R pour Répétable ; F pour Facultatif.

Grâce au format ONIX, la BnF et tous ses partenaires ont une façon commune de gérer les métadonnées. Le fichier ONIX décrivant un livre numérique est obligatoire pour le dépôt légal de cet ouvrage. Les fichiers de métadonnées et ceux des livres numériques seront chargés sur la plateforme sFTP par le distributeur. Le traitement automatique commence par des vérifications (présence des deux fichiers, contrôles des EPUB et des PDF, validation des fichiers ONIX) et, une fois que les fichiers sont certifiés conformes et adaptés aux demandes de la BnF, les métadonnées sont converties pour l'étape suivante, le catalogage.

Organisation du circuit de catalogage

Le but de cet accord avec les éditeurs est de construire un circuit qui permette aux métadonnées d'intégrer automatiquement la chaîne de catalogage de la bibliothèque. La BnF avait déjà une expérience de la gestion des fichiers ONIX décrivant les livres imprimés, comme il a été dit plus tôt. Ce dernier travail reste, d'un point de vue intellectuel, proche de la méthode traditionnelle : les livres imprimés arrivent de la manière habituelle et sont pris en main par les catalogueurs.

Le circuit du livre numérique est assez similaire en termes de traitement des métadonnées. Le fichier ONIX, qu'il décrive un livre imprimé ou un livre numérique, est converti en notice INTERMARC de la même façon. Le résultat de cette conversion est une notice bibliographique de base relativement complète, puisqu'elle mentionne déjà la majorité des informations (titre, contributeurs, éditeur...). Les catalogueurs doivent ensuite vérifier les informations et enrichir la notice avec les autorités, l'indexation et la description bibliographique correcte pour atteindre le niveau de référence.

L'organisation du catalogage des livres numériques est encore en phase de préfiguration, mais la majorité des éléments est déjà en place. Durant la phase de développement du prototype, une équipe composée de cinq catalogueurs et d'une coordinatrice va prendre en charge ce catalogage. Rejoignant la bibliothèque entre juin et décembre 2015, ils feront partie intégrante du service qui catalogue déjà tous les livres imprimés arrivant au dépôt légal. Cette équipe sera capable de cataloguer des livres imprimés aussi bien que des livres numériques ; elle sera d'abord seule à cataloguer les livres numériques, mais le but est de diffuser cette expertise et de faire en sorte que chaque catalogueur du service soit capable de traiter les livres numériques.

Néanmoins, commencer avec une petite équipe nous permettra d'étudier et d'expérimenter une nouvelle façon de travailler : de quel équipement ces catalogueurs ont-ils besoin (deux écrans, par exemple) ? Comment organiser leur formation ? Étant donné que tout le circuit est automatisé, les catalogueurs ne peuvent avoir accès aux livres numériques que lorsqu'ils sont disponibles dans la bibliothèque numérique, après l'étape de préservation⁵. Leur travail sur la notice intervient donc vraiment en bout de chaîne et clôt le traitement. Ainsi, les catalogueurs auront la possibilité de cataloguer « livre en main » – ou, plus précisément, sous les yeux. Cependant, des questions demeurent sans réponse, notamment celle de l'adéquation de l'interface publique de la bibliothèque numérique à une consultation pour catalogage.

À côté de ces points très pragmatiques, nous avons dû réfléchir à un sujet presque conceptuel : la définition d'une notice décrivant un livre numérique. Cette définition a

⁵. Sur le circuit en général, voir [1].

déterminé le schéma de conversion des métadonnées des éditeurs en métadonnées de la bibliothèque et donc les informations que les catalogueurs auront à traiter.

À quoi pourrait ressembler une notice de livre numérique dans le catalogue de la BnF ?

Pour répondre à cette question, un groupe de travail sur les aspects bibliographiques a été spécialement créé au début du projet, afin d'étudier les besoins de la chaîne du livre numérique en la matière. Établir le contexte bibliographique a en effet été utile pour étayer d'autres aspects du projet (par exemple les ressources humaines pour le catalogage, ou la préservation). Ce groupe de travail se réunit encore régulièrement pour suivre les évolutions sur ce sujet.

Les personnes participant à ce groupe de travail reflètent la complexité de la nature des livres numériques : ils sont membres du département du Dépôt légal (chef du projet et département dans lequel sont catalogués les livres déposés), du département de l'Audiovisuel (qui prend soin du dépôt légal des documents multimédias) et du département de l'Information bibliographique et numérique (en charge des métadonnées et du respect des normes).

Pour définir ce que pourrait être une notice de livre numérique, ce groupe a dû se confronter à des questions ontologiques : la description bibliographique d'un livre numérique doit-elle être plus proche de celle d'une monographie ou d'un document électronique ? Le fait que ces livres numériques soient traités par des catalogueurs qui ont affaire à des monographies imprimées constitue un élément de contexte important. Pour le moment, la décision a été prise de traiter ces documents davantage comme des livres que comme des documents électroniques.

Une fois ces choix tranchés, la question suivante était de savoir comment intégrer de nouvelles métadonnées dans le modèle BnF. D'un côté, certaines informations n'existent pas encore à l'intérieur du Catalogue général de la bibliothèque (un format de fichier et ses versions successives, par exemple). D'un autre côté, des usages bibliographiques existent déjà pour décrire les collections très diverses de la BnF et ils doivent être pris en compte : par exemple, la notion de poids du document est déclarée dans un champ spécifique de notre format, mais ce champ est actuellement utilisé pour enregistrer le poids d'une pièce ou d'une médaille, un autre champ doit donc être trouvé pour le « poids » d'un fichier (sa taille).

Dans le même temps, la réflexion sur la notice idéale de livre numérique permet des innovations qui constituent une force motrice pour les livres imprimés. Par exemple, en travaillant avec les éditeurs sur le modèle de métadonnées ONIX, nous leur avons demandé de nous fournir le code-barres EAN d'autres versions du même livre, un champ qui existe dans ONIX (<RelatedProduct>). Grâce à cette information, un lien est créé entre les différentes notices durant la conversion automatique de l'ONIX en notice MARC, si le catalogue contient déjà une notice avec ledit EAN. La même chose sera possible avec des notices d'autorité auteur, grâce à l'ISNI. En utilisant les liens entre notices du même livre dans différents formats, il sera alors possible d'extraire toutes les informations utiles de la notice la plus complète pour enrichir automatiquement les autres. Selon la version du livre qui est cataloguée en premier, cela signifie non seulement qu'une notice créée pour un livre imprimé peut être utilisée pour enrichir une notice de PDF ou d'EPUB, mais que l'inverse sera aussi possible.

Par conséquent, des notices de différents niveaux de complétude et de conformité aux normes coexisteront dans le catalogue. Une telle situation est fréquente, mais celle-ci sera particulièrement complexe, car il ne s'agira pas d'une simple différence de notice entre le

livre imprimé et le livre numérique. Certains livres numériques auront des notices plus riches du fait de l'extraction des métadonnées d'un livre imprimé préexistant, le contraire pourra aussi arriver. Ainsi, le processus de traitement lui-même induira une telle symbiose entre les notices que nous ne serons pas capables (même si nous le voulions) de séparer la chaîne du livre imprimé de celle du livre numérique.

Le recoupement existe aussi en termes de normes : la norme française applicable aux livres numériques est la NF Z 44-082 *Catalogage des documents électroniques*, qui date de 1999. Elle donne seulement des indications sur l'enregistrement des caractéristiques propres aux documents électroniques, le reste de la ressource étant décrit selon les spécifications de la norme correspondante (livres imprimés, cartes, périodiques, etc.). Bien sûr, cette norme, conçue avant l'essor des livres numériques, est maintenant obsolète. C'est un autre point à prendre en compte pour la conception du schéma de catalogage des livres numériques.

« Dans le catalogue et au-delà » [2]

Notre première réaction face à la question « comment fait-on entrer les livres numériques dans le catalogue ? » s'est traduite en termes de format, car, très concrètement, c'est ainsi que nous cataloguons. Mais la réponse se trouve dans les normes.

Il n'existe pas d'ensemble de normes spécifiquement dédié aux livres numériques en France aujourd'hui. Les règles dérivent de celles qui s'appliquent aux livres et aux documents électroniques, qu'il s'agisse des règles de catalogage françaises ou de l'ISBD. Mettre à jour les règles de catalogage françaises représente un enjeu plus vaste que celui du seul livre numérique, en particulier dans le contexte global de transition vers RDA. La France a mis en place un groupe de travail sur l'adoption de ces règles de catalogage depuis 2010. Fin 2014, les deux agences bibliographiques nationales ont pris la décision d'une politique de convergence vers RDA : adopter autant du code que possible, sans mettre en péril nos pratiques nationales lorsqu'elles sont plus riches ou plus conformes à FRBR que RDA ne l'est actuellement.

Sous le nom de « transition bibliographique », le groupe est à présent chargé de mettre à jour les normes françaises sous la forme de ce que nous appelons RDA-FR. La procédure étant très longue, des parties de la nouvelle norme sont publiées et remplacent progressivement les parties correspondantes de la norme française. Cela signifie que nous travaillons actuellement avec deux normes à la fois. Mais l'approche intégrée de RDA, avec des règles générales qui concernent tous les types de documents, pourrait être utile à la convergence des notices.

La nature du livre numérique est un obstacle important pour les bibliothécaires. Nous avons tendance à les considérer d'abord et avant tout comme des livres et nous sommes tentés de les traiter comme le dépôt légal imprimé : créer des notices de référence que nous publions ensuite dans la Bibliographie nationale.

Cependant, ce n'est pas ce que nous dit la loi. Les livres numériques entrent dans le cadre de la loi sur le dépôt légal numérique, qui a été créée en pensant au web. Cela signifie que la Bibliothèque nationale n'a aucune obligation d'instituer un dépôt légal exhaustif pour le livre numérique, ni de publier une bibliographie nationale. Mais toute notre réflexion sur les livres numériques nous a conduits à mettre l'accent sur les points communs plutôt que sur les différences entre livre imprimé et livre numérique, et à créer un modèle de description plus proche du livre imprimé que du web.

En outre, nous savons que le besoin existe d'une base de données ou d'un registre national détaillé pour les livres numériques, avec des notices de référence fiables : il a été exprimé

dans le rapport Lescure à la ministre de la Culture en 2013 [3]. La base de données commerciale Electre a récemment commencé à fournir des métadonnées pour les livres numériques. Mais la BnF a une position unique d'observatrice du monde de l'édition imprimée et numérique. De plus, le principe même de produire des notices individuelles pour les livres numériques dans le catalogue de la bibliothèque semble être un argument en faveur du passage à l'étape suivante, leur publication dans la Bibliographie nationale française.

La question, bien sûr, est de savoir comment.

En termes de qualité, la Bibliothèque nationale est une fournisseuse de service pour la communauté nationale, notamment à travers la Bibliographie nationale qui, contrairement au Catalogue général, contient traditionnellement uniquement les notices de la plus haute qualité. Mais les *Guidelines for bibliographies in the electronic age (Les bibliographies nationales à l'ère du numérique)* [4] ont identifié le problème et admis que, pour les livres numériques, il est possible d'avoir différents niveaux de description. Il nous incombe de déterminer quel niveau de description adopter, en gardant à l'esprit que notre processus rend les notices de livres numériques et de livres imprimés étroitement interdépendantes.

En termes d'affichage, faut-il afficher les notices dans le volet *Livres* de la bibliographie, mélangées à celles des livres imprimés, avec simplement un index spécifique ? En termes de développements informatiques, ce serait la solution la plus simple et la plus rapide. Cela pourrait être une solution temporaire, jusqu'à ce qu'un nouveau volet de la bibliographie ou un nouveau site web puisse être créé pour l'ensemble de la publication. Si nous décidons de mettre à jour le site web actuel, quelle forme devrait-il prendre ? Lorsque nous aurons éventuellement plusieurs versions d'un même titre, un affichage « FRBRisé » pourrait être envisagé. Examiner les autres sites web que la BnF utilise pour afficher les métadonnées constituerait la première étape de cette étude.

Conclusion

Une fois dépassé le premier réflexe, qui consiste à faire ce que nous avons « toujours » fait, nous devons nous demander quel est le service qui est attendu de nous. Avec les livres numériques, le contexte du dépôt légal change : les livres numériques portent avec eux des métadonnées formatées et normalisées. Même si les normes ne sont pas celles que nous, bibliothécaires, utilisons, nous recevons des notices qui ne sont plus des notices de base et qui offrent déjà un service aux usagers. Mais est-ce le service que nous souhaitons offrir ? À ce stade, cela paraît un grand pas dans la bonne direction, mais du chemin reste à parcourir pour atteindre le niveau prescrit par les normes. Comment combler cet écart ? L'automatisation, qui permettra d'importer les données afin d'avoir le même niveau de description entre les différentes notices d'un même titre, sera d'une grande aide, mais elle a ses limites. C'est à ce stade que l'intervention humaine devient nécessaire.

Ce projet met en valeur la convergence des questions auxquelles sont confrontés les bibliothécaires : l'adoption de RDA, la révision des normes, l'automatisation du catalogage. Pour répondre à ces questions, nous devons avoir une vue plus globale de notre activité et transcender la structure parfois rigide de la bibliothèque.

Références

- [1] S. Derrot et C. Oury, “Ebooks: rather electronic or book? Extending legal deposit to ebooks at the Bibliothèque nationale de France”, in IFLA WLIC 2014, 16-22 août 2014, Lyon, France. [<http://library.ifla.org/830/>]
- [2] International Federation of Library Associations and Institutions, Tillett, B. B., et Cristán, A. L., *IFLA cataloguing principles: The Statement of International Cataloguing Principles (ICP) and its glossary : in 20 languages*, Munich, K.G. Saur, 2009. [http://www.ifla.org/files/assets/cataloguing/icp/icp_2009-en.pdf]
- [3] Pierre Lescure, *Contributions aux politiques culturelles à l'ère numérique*. [Rapport à la ministre de la Culture], 2013. [<http://www.ladocumentationfrancaise.fr/rapports-publics/134000278/index.shtml>]
- [4] M. Žumer (ed.), *National bibliographies in the digital age: guidance and new directions*, Munich, Saur, 2009. [<http://public.eblib.com/choice/publicfullrecord.aspx?p=454007>]