

CHAPITRE DE LIVRE :

« Le livrel et le format ePub »

Fabrice Marcoux

dans *Pratiques de l'édition numérique*, Montréal, Presses de l'Université de Montréal, coll. « Parcours numérique », 2014, p. 177-189.

Pour citer ce chapitre :

MARCOUX, Fabrice, « Quelles actions sur le web ? », dans Michael E. Sinatra et Marcello Vitali-Rosati (dir.), *Pratiques de l'édition numérique*, Montréal, Presses de l'Université de Montréal, coll. « Parcours numérique », 2014, p. 177-189.

CHAPITRE 11

Le livrel et le format ePub

FABRICE MARCOUX

Il ne serait pas possible de tracer un panorama complet des enjeux de l'édition numérique sans parler des formats qui tentent de reproduire le mode typique de circulation des contenus papier – le livre – en l'adaptant au support numérique. C'est ce que l'on appelle «livre électronique» ou «livrel» (*eBook* en anglais). Il ne faut pas confondre le livre électronique avec la liseuse qui est le support de lecture. Le livre électronique est un fichier, formaté selon des standards déterminés. Bien évidemment, ces formats ne sont pas stables et changent très rapidement. Aujourd'hui, le standard ouvert de référence est l'ePub. Ce chapitre essaie d'en donner une description.

Introduction au livre numérique

Le *Grand dictionnaire terminologique* de l'Office québécois de la langue française définit ainsi le livre numérique, aussi appelé « livrel » : « Livre disponible en version numérique, sous forme de fichier, qui peut être téléchargé, stocké et lu sur tout appareil électronique qui en permet l'affichage et la lecture sur écran ».

Cette définition est assez générale au vu du fait qu'il existe plusieurs types de livrels : on peut notamment en identifier trois. Le premier, et le plus répandu, est le livrel « homothétique », qui est une transposition à l'identique d'un livre papier en version numérique. Le deuxième est le livrel enrichi, qui prend une place de plus en plus importante : il utilise les possibilités techniques du format numérique afin d'apporter un enrichissement, autant au contenu qu'à la mise en forme de l'ouvrage imprimé qu'il vient compléter. Enfin, il existe un troisième type de livrels : le livrel « originairement numérique », créé *par et pour* le numérique, et non en complément ou par imitation d'un « original papier ».

Le livrel a commencé à prendre de l'importance sur le marché à partir de 2003 et, depuis 2010 environ, les nouveautés publiées par la majorité des maisons d'édition sont immédiatement disponibles en version « homothétique ». Le prix des livrels avoisine généralement 75 % du coût de la version papier et peut s'élever à 100 % (dans le cas des livres présentant de nombreuses illustrations, par exemple). Les livrels sont souvent protégés contre la copie par filigrane et parfois par verrou numérique (DRM, *Digital Rights Management*).

Parmi les formats existants qui permettent de réaliser un livrel, le plus populaire est actuellement l'ePub, format

de fichiers non propriétaire maintenu par l'International Digital Publishing Forum (IDPF) qui a pour mandat d'en faire le standard pour l'édition de livre numérique. L'ePub est une norme ouverte qui permet de créer des livrels inspirés du web ou de livres papier, ou encore de faire des versions enrichies de livres papier pour les liseuses électroniques et pour le web.

Historique de l'ePub

L'ePub a pour ancêtre le format Open eBook, créé par SoftBook Press, société fondée en 1996. La norme *Open eBook Publication Structure* (OEBPS), à la base de la norme OPS (*Open Publication Structure*), fut élaborée à partir de la technologie développée pour la liseuse électronique SoftBook, lancée en 1998.

La prolifération des formats, survenue en 1998-1999, a créé un éclatement de l'offre. Chaque format de livrel n'est alors compatible qu'avec un seul modèle de liseuse : la nécessité de mettre en place une norme ouverte et commune s'impose.

Dès 1998, le National Institute of Standards and Technology (NIST) initie le processus de normalisation en mettant sur pied l'Open eBook Initiative. C'est ce groupe qui «élabore l'*Open eBook* (OEB), un format de livres numériques basé sur le langage XML (*eXtensible Markup Language*: langage de balisage extensible) et destiné à normaliser le contenu, la structure et la présentation des livres numériques¹ ». En septembre 1999, la version 1.0 de l'OEBPS, sur laquelle se fonde le format OEB, est déjà disponible.

1. Marie Lebert, *Une courte histoire de l'eBook*, NEF, Université de Toronto, 2009, p. 75.

En janvier 2000, l'Open eBook Initiative devient l'Open eBook Forum puis, en avril 2005, l'International Digital Publishing Forum (IDPF). Ce consortium a alors une double vocation de commercialisation et de normalisation. Sa mission centrale est « d'établir une norme globale, interopérable et accessible pour les livres électroniques et d'autres publications, afin de contribuer à la croissance de l'industrie de l'édition numérique² ». Quelques mois plus tard, en juillet 2005, le format ePub remplace l'OEB. Mais ce n'est qu'en 2007 que l'ePub2 devient une norme de l'IDPF.

L'IDPF lance, en octobre 2011, l'ePub3 : le format ePub s'appuie dès lors sur les règles du HTML5 (plutôt que sur le XHTML) pour la structuration du contenu et sur celles des feuilles de style CSS3 pour la mise en forme.

En mars 2013, l'IPA (Union internationale des éditeurs) déclare officiellement qu'elle reconnaît le format ePub3 comme norme internationale. Au moment d'écrire ces lignes, l'ePub3 suit le processus d'accréditation pour recevoir le statut de norme ISO/IEC (International Organization for Standardization/International Electrotechnical Commission).

Particularités techniques du format ePub

L'ePub est basé sur les mêmes langages de balisage que ceux employés pour la réalisation de sites web : il s'agit de fichiers HTML. C'est donc un format permettant de faire des livres numériques ayant à la fois les caractéristiques du livre papier et les caractéristiques d'un site web.

2. Bill McCoy, « Portable Documents for the Open Web – Part 1: What Role does ePub Play in the Cloud-Centric World? », *O'Reilly-TOC*, 2012 (traduction libre). Bill McCoy est président de l'IDPF.

Le fichier de format ePub est organisé selon une arborescence qui comprend un dossier dans lequel figurent des fichiers HTML, des CSS et des documents de différents types (audio, image, vidéo, etc.) regroupés eux-mêmes par sous-dossiers. Ce qui procure à cet ensemble une unité « organique » est la rigueur avec laquelle tous les éléments sont répertoriés dans l'élément « *manifest* » du fichier OPF (document structuré fondé sur XML). Il faut également que tous les documents du contenu soient regroupés dans le dossier OPS. Le dossier initial est ensuite compressé suivant le protocole d'archivage ZIP (norme ouverte de compression de dossiers). L'extension « .epub » vient remplacer l'extension « .zip » du fichier qui en résulte.

Le principe du « bien formé » qui préside à la construction des documents structurés fondés sur XML suppose qu'il ne doit y avoir qu'un seul élément racine dans lequel tous les autres s'emboîtent. Un document XML est nécessairement du même type que son élément racine (<html> dans le cas des pages web). Mais, pour éliminer toute ambiguïté quant à ce que signifie ce type, il peut être opportun d'introduire, au début du document, une déclaration XML qui renvoie à une DTD (définition de type de document). La DTD sert à prescrire ce que peut contenir chaque élément et sous-élément. Tout élément de contenu doit être encadré par une balise ouvrante (<élément) et une balise fermante (élément/>). Les sous-éléments doivent être entièrement compris dans les éléments de niveau supérieur (pas de chevauchements). Les consignes de mise en forme, s'il y a lieu, sont complètement séparées du contenu. On peut ainsi changer la mise en forme sans affecter le contenu, ou

extraire des informations du contenu sans que le code des consignes de mise en forme fasse interférence.

Le détail des spécifications techniques constitutives des deux versions du format (ePub2 et ePub3) est disponible gratuitement sur le site de l'IDPF. Celles-ci sont formelles et indiquent ce sur quoi doivent se baser les développeurs qui veulent créer une application capable de lire correctement un fichier au format ePub. Il est possible de faire valider un fichier réalisé suivant l'une ou l'autre des versions de la norme recommandée par le consortium.

Caractéristiques générales

Plusieurs des caractéristiques générales de l'ePub sont reliées à l'adoption du principe des documents structurés.

La première caractéristique de l'ePub est de pouvoir contenir tous les éléments typiques d'un document sur le web. En d'autres mots, le format ePub est un cousin des sites web et leurs structures sont semblables. Les documents contenus dans un fichier ePub sont rangés dans les dossiers pour les chapitres (texte), les images, le son (audio) et les films (vidéo). Les deux versions supportent bien les hyperliens. La version 3.0, à l'instar du HTML5, permet les scripts (Javascript, une autre norme ouverte) dans les documents de contenus.

La deuxième caractéristique de l'ePub est de permettre d'avoir l'ensemble des éléments paratextuels propres au livre : un livrel peut avoir une table des matières, un index, une page de couverture. Un nom de fichier est réservé pour chacun de ces éléments. Enfin, ce qui rapproche beaucoup l'ePub du livre est le principe de répartition du contenu : un chapitre par fichier. Cependant, les fonctionnalités typiques

du numérique permettent d'aller au-delà de la structuration linéaire du livre papier. À travers des systèmes de signets, on permet de transformer une idée héritée du livre papier en une véritable forme d'hyperliens.

Les livrels ePub (2 ou 3) peuvent contenir divers éléments caractéristiques des deux environnements (web et livre papier). Ceux-ci incluent notamment les tableaux, les listes, les images (et autres figures) et les encadrés (<div>). S'y ajoutent les divisions qui servent pour l'en-tête et le pied de page, et qui sont explicitement prévues dans le HTML5. Il en va de même pour les « notes de bas de page », qui deviennent des « notes marginales » (élément « *aside* »). Des graphiques de type SVG (*Scalable Vector Graphics* – Graphiques vectoriels adaptables), basés sur le balisage XML, peuvent également être inclus, garantissant ainsi l'interopérabilité. Il s'agit de dessins de type vectoriel, qui se construisent avec des balises et leurs attributs (du texte) et non par du code binaire.

À ces caractéristiques de base du fichier ePub peuvent se greffer des dispositifs optionnels tels que les dictionnaires et les DRM.

Certaines applications, comme Aldiko, offrent la possibilité de consulter des dictionnaires en cliquant sur des mots contenus dans le livrel. L'utilisateur peut ainsi avoir la définition d'un terme en temps réel, et éventuellement sa traduction en une autre langue. Ces dictionnaires peuvent être inclus dans l'application de lecture, ou alors être disponibles sur le web et demander une connexion pour être consultés.

En ce qui concerne les DRM, dans le cadre d'une « édition homothétique », les éditeurs peuvent avoir tendance à essayer de transposer en numérique le modèle économique

de l'édition papier. Ce modèle est cependant mis en crise par la facilité de copier des livres numériques. De plus, les versions « homothétiques » livrent l'intégralité du contenu du livre papier. Les concepteurs du format ePub ont alors prévu la possibilité de configurer des verrous numériques (DRM), pensant qu'une majorité d'éditeurs n'adopteraient ce format qu'à cette seule condition. Mais ces restrictions au partage n'étant pas unanimement acceptées, l'IDPF a rendu les DRM optionnelles.

Fonctionnalités

Passons maintenant en revue les six principales fonctionnalités de l'ePub (et des autres livrels).

Premièrement, il est portable puisqu'en tant que fichier informatique il ne pèse rien de plus physiquement par rapport à l'appareil de lecture (sur lequel on peut stocker des milliers de livrels). Et son contenu est « recomposable » (*reflowable*), le rendant consultable sur des plateformes variées. Le texte flottant, à savoir le fait que le texte se met en page automatiquement selon la taille de l'écran et les options choisies, est un des traits qui distinguent l'ePub du PDF.

Deuxièmement, l'ePub permet la recherche en plein texte. L'architecture de l'information rigoureuse qui le caractérise donne la possibilité de créer des applications pour aller chercher (extraire) des données. Dans le cas de l'ePub3, les CFI (*Canonical Fragment Identifiers*) améliorent cette capacité. Il s'agit d'une spécification qui définit des méthodes standards pour baliser puis référencer des fragments de contenus (un mot, une partie du texte, une image, etc.) d'un fichier ePub. Cela permet de créer des hyperliens

et de naviguer plus aisément, de façon non linéaire, dans les contenus d'un fichier.

Une troisième fonction fondamentale des documents ePub est la possibilité d'associer des annotations au contenu d'un livrel. Il faut préciser que cette fonctionnalité dépend pour le moment davantage de l'application que du format. C'est pourquoi l'ePub3 permettra d'associer des annotations plus riches grâce, justement, aux identifiants de fragments (CFI). Ces identifiants devraient aussi favoriser le partage des annotations et des marque-pages (ou signets) qui sont une forme d'annotation standard dans la plupart des logiciels de lecture.

L'ePub3 offre davantage de possibilités d'« échanges » avec le livrel que l'ePub2. Il prévoit ainsi la création de quiz, des animations dans les fenêtres surgissantes (*pop-up*), des couches de multimédias superposées, etc. Les effets de styles autorisés par l'ePub 3 (Javascript et CSS3) sont également plus riches, mais l'IDPF recommande d'utiliser animations et autres scripts « sophistiqués » avec « prudence ».

En quatrième lieu, les fichiers ePub sont configurables. Pour les options de présentation, ils permettent un ajustement des paramètres concernant le rendu du texte à l'écran (polices, tailles, marges). Il est aussi possible de déterminer une couleur de fond d'écran, pour atténuer au besoin le contraste, par exemple. D'autres aspects peuvent être configurés, comme la prise en charge des jeux de caractères particuliers, requérant parfois le chargement de bibliothèques, comme MathML. Ce langage pour l'affichage correct de formules mathématiques est supporté par l'ePub3, grâce à HTML5 et CSS3.

Il est aussi faisable – et même souhaitable parfois pour les livrels enrichis – de configurer le livrel avec une mise en page déterminée (*fixed layout*), même lorsqu'il s'agit d'un format ePub (version 3). Mais on perd alors l'avantage de la mise en page « flottante » (« recomposable »).

En cinquième lieu, les applications de lecture proposent généralement plusieurs options pour la présentation des pages et la manière de circuler à travers le contenu. Les flèches à gauche et à droite permettent de revenir en arrière ou d'avancer. Une barre de défilement permet de se situer dans l'ensemble du document et de se rendre directement à un autre point. Un champ présentant le numéro de page peut être visible et offrir la possibilité d'entrer un autre numéro pour accéder directement à la page choisie (quand les pages sont indexées). On peut compter sur la présence d'une table des matières avec hyperliens vers les parties en question (pourvu que les parties du contenu aient été incluses dans le fichier assigné à cet usage : le fichier TOC). L'index est un autre moyen de se repérer pour naviguer.

Sixième fonctionnalité fondamentale, les métadonnées globales sont incluses dans l'élément « *metadata* » du fichier dont l'extension est « .opf », et sont structurées en Dublin Core. Le Dublin Core est une norme pour les métadonnées du milieu documentaire. Pour référencer des fragments au moyen d'un identifiant numérique (dans le but de pouvoir y associer des métadonnées), l'ePub3 spécifie comment procéder grâce au « protocole epubcfi ». Les métadonnées permettent d'évaluer la pertinence d'un document et de le qualifier. Elles sont « encapsulées » dans l'enveloppe (l'élément « *package* ») du document.

Évaluation

Les possibilités énumérées ci-dessus découlent de la structure même du format ePub : il est nécessaire que les concepteurs de chaque livrel les implémentent de manière optimale pour en tirer tout le bénéfice. Il faut néanmoins tenir compte des limites des applications de lecture, dont l'ergonomie peut rendre parfois difficile l'exploitation des fonctionnalités offertes par ce format.

On peut proposer un bilan provisoire du développement du format ePub en essayant d'en montrer les principales qualités mais aussi d'en identifier les limites.

On constate que la plupart des forces principales de l'ePub proviennent de son organisation et de la cohérence que lui procure l'utilisation des principes de séparation de la forme et du contenu, caractéristiques des documents structurés.

La première chose qu'un auteur souhaite, quand il confie son livre à un éditeur, est que son œuvre soit lue. Or, si le fait de publier par et pour le web ouvre à tout un nouveau public, il faut relever le défi de rendre les contenus visibles à travers la mer d'informations qui envahit la toile.

Le fait que les fichiers électroniques permettent la recherche en plein texte est un avantage. Mais s'ils ne sont pas repérés eux-mêmes, c'est totalement inutile. Or l'ePub présente l'avantage, justement, de favoriser la repérabilité de ses contenus, puisqu'il en présente (s'il est bien fait) les informations pertinentes sous forme de métadonnées globales (Dublin Core) et spécifiques (CFI). La possibilité de bien structurer les métadonnées dans le fichier ePub est tout à fait alignée au développement progressif du web vers le web sémantique. En ce sens, il n'en tient qu'aux éditeurs

de profiter de l'occasion que l'ePub leur offre de rendre leurs contenus beaucoup plus repérables et récupérables en s'assurant d'indiquer clairement dans l'élément « *metadata* » toutes les informations utiles.

Avec l'ePub3, il est aussi possible d'entrer plus profondément dans la structure pour qualifier des fragments de documents (CFI) qui méritent d'être portés à l'attention des moteurs de recherche. L'attribution des métadonnées est un savoir-faire qui doit être appris et appliqué.

L'accessibilité est d'une importance capitale dans le contexte de la société de l'information. Cela passe par l'adoption de normes communes par les développeurs de navigateurs (W3C). À l'instar des pages web, les livrels deviennent des outils pour relier des contenus de types différents. Il est donc important que des normes assurant la conformité à certains standards soient établies et suivies. C'est ce qui donne la possibilité aux créateurs d'applications d'assistance technique de fournir des outils universellement applicables pour les personnes aux prises avec des obstacles à la lecture ou à l'écoute. Ces extensions des logiciels de lecture rendent le contenu disponible autrement pour pallier une difficulté d'accessibilité. LePub se positionne avantageusement à cet égard depuis longtemps.

La grande force du format ePub est certainement d'être ouvert. Comme nous l'avons indiqué, cela signifie que les spécifications constitutives du format sont disponibles gratuitement pour tout le monde. De cette façon, il est possible à tous les distributeurs, éditeurs ou développeurs de se baser sur ces indications pour élaborer des logiciels qui seront capables de lire les livrels au format ePub. Il n'y a ainsi pas de raison pour qu'une plateforme populaire

n'offre pas la possibilité de lire des ePub. C'est ce que l'on appelle l'interopérabilité. Et, surtout, cela encourage la collaboration.

On peut, en revanche, indiquer certaines limites du format ePub.

Premièrement, il faut reconnaître que les contraintes liées au langage de balisage peuvent rendre rigide la structure qui devra être donnée aux livrels pour qu'ils puissent être validés. Deuxièmement, même si les principes du XML qui sous-tendent l'organisation des livrels au format ePub sont d'une logique relativement simple à comprendre, ils supposent la connaissance de règles syntaxiques rigoureuses et la maîtrise de nouveaux langages qu'il peut être difficile – et coûteux pour les éditeurs – d'apprendre. Troisièmement, les deux versions de la norme (ePub2 et ePub3) peuvent, pour le moment, apporter de la confusion.

En dernier lieu, il pourrait y avoir des résistances à adopter le format ePub en raison de l'attrait supplémentaire qui offrent des formats concurrents. Les formats privatifs, en particulier, appartenant à des corporations comme Amazon (AZW et KF8) et Apple (iBooks, un format privatif dérivé de l'ePub), bénéficient de la visibilité que leur offrent les plateformes de distribution et les appareils de lecture de ces compagnies.