

Référentiel ePub 3



Source gallica.bnf.fr / Bibliothèque nationale de France

<http://gallica.bnf.fr/ark:/12148/btv1b6929540f.r>

Bibliothèque nationale de France direction des Services et des réseaux département de la Conservation service Numérisation	Date :le mardi 29 octobre 2013 Version :1 Référence BnF :BnF-ADM-2013-078731-01
---	---

TABLE DES MATIERES

1.	INTRODUCTION	6
1.1	DOCUMENTS APPLICABLES ET DE REFERENCE	6
2.	INTRODUCTION AU FORMAT EPUB	8
3.	PROCESSUS DE PRODUCTION EPUB	9
3.1	REFECTION DE DOCUMENTS DEJA NUMERISES	9
3.1.1	OCR en qualité éditoriale	9
3.2	PRODUCTION CONCOMITANTE A UN MARCHE DE NUMERISATION	9
3.2.1	OCR en qualité éditoriale	9
4.	CHOIX DES DOCUMENTS A TRAITER EN EPUB	10
4.1	CRITERES TECHNIQUES	10
4.1.1	Origine des documents	10
4.1.2	Qualité OCR des documents	10
4.2	CRITERES BIBLIOGRAPHIQUES	12
4.3	CRITERES INTELLECTUELS	13
4.4	RECOMMANDATIONS GENERALES	13
4.4.1	Mode de sélection	13
4.4.2	Avertissement éditorial	14
4.4.3	Contrôle de la sélection	14
5.	CONVERSION VERS LE FORMAT EPUB	15
5.1	MONTEE EN QUALITE DES CONTENUS TEXTE	15
5.1.1	Périmètre de la montée en qualité du texte	15
5.1.2	Erreurs majeures	16
5.1.3	Erreurs mineures	17
5.1.4	Valeurs des taux qualité	17
5.2	STRUCTURATION	18
5.2.1	Périmètre de la montée en qualité	18
5.2.2	Segmentation	18
5.2.3	Structuration logique	18
5.3	MISE EN FORME	24
5.3.1	Styles de paragraphe	24
5.3.2	Styles de caractère et typographie	28

5.3.3	Illustrations	29
5.3.4	Éléments graphiques	34
5.3.5	Tableaux	39
5.3.6	Formules	39
5.3.7	Publicités	40
5.3.8	Notes	40
5.3.9	Références croisées	43
5.3.10	Divers	44
5.4	COUVERTURE ET PAGE DE TITRE	45
5.4.1	Couverture	45
5.4.2	Page de titre	50
5.5	TABLES DE NAVIGATION	51
5.5.1	Table des matières	51
5.5.2	Table des repères	54
5.5.3	Table des numéros de pages	55
5.6	INDEX	57
5.7	PAGES SPECIALES DE L'ORIGINAL	57
5.7.1	Reliures	57
5.7.2	Achevés d'imprimé, mentions d'imprimeur, etc.	57
5.7.3	Publicités	58
5.7.4	Errata	58
5.7.5	Tables	58
6.	METADONNEES ET MENTIONS UTILES	60
6.1	METADONNEES BIBLIOGRAPHIQUES	60
6.1.1	Identifiant	60
6.1.2	Titre	61
6.1.3	Auteur	64
6.1.4	Contributeurs	65
6.1.5	Editeur	65
6.1.6	Dates	66
6.1.7	Langue	67
6.1.8	Source	67
6.1.9	Relation	68
6.1.10	Coverage	68
6.1.11	Sujet	68
6.1.12	Type	68

6.1.13	Droits	69
6.1.14	Description	69
6.2	MENTIONS UTILES	69
6.2.1	Page « À propos de cette édition numérique »	69
7.	CARACTERISTIQUES TECHNIQUES	72
7.1	VERSION ET PROFIL	72
7.1.1	Livre numérique et accessibilité	72
7.1.2	Variante <i>fixed layout</i>	73
7.2	FEUILLES DE STYLES CSS	74
7.3	DECOUPAGE DES CONTENUS	74
7.4	POLICES DE CARACTERES	74
7.5	IMAGES	74
7.6	LIVRAISON	75
7.6.1	Formats	77
8.	REJET DES DOCUMENTS A TRAITER EN EPUB	78
8.1	CAS DES EPUB « AMENAGES »	78
9.	CONTROLES APPLIQUES PAR LA BNF	79
9.1	CONTROLES AUTOMATIQUES	79
9.1.1	Contrôle technique	79
9.1.2	Contrôle de l'ark	79
9.1.3	Contrôle du format général de l'ePub	79
9.1.4	Contrôle de structure général de l'Epub	80
9.2	CONTROLES VISUELS	84
9.2.1	Défauts majeurs	85
9.2.2	Défauts mineurs	88
10.	TERMINOLOGIE	89
10.1	GLOSSAIRE	89
10.2	ABRÉVIATIONS	90

1. INTRODUCTION

La BnF a récemment décidé d'accroître son effort de valorisation des contenus patrimoniaux sous la forme de livres numériques au format ePub. Cette nouvelle forme de valorisation vise à tirer parti des atouts du livre numérique comparé aux formats classiques proposés par les bibliothèques numériques (web, PDF) :

- lecture nomade, hors site web de la bibliothèque numérique,
- lecture sur un dispositif dédié, dans un format dédié (ePub),
- dissémination des contenus facilitée.

Du fait des caractéristiques propres du livre numérique et de ses formats techniques, cette action est pensée en tant que diversification des formats de consultation de la bibliothèque numérique. Elle n'a pas d'objectif de conservation. En effet, la conversion d'un ouvrage en livre numérique implique le plus souvent la perte d'informations de forme (mise en page, typographie, etc.), voire de contenus.



Pourquoi le format ePub et non le PDF ?

La BnF a préféré le format ePub au format PDF comme support de l'action de publication de livres numériques car il s'agit d'un format recomposable : contrairement à un format paginé comme le PDF, un livre numérique au format ePub a la capacité de s'adapter à tout type de périphérique ou d'écran de lecture. Cette caractéristique en fait un format particulièrement adapté aux contenus majoritairement textuels.

1.1 Documents applicables et de référence

Norme ePub version 3.0	http://idpf.org/epub/30
Spécification pour la publication	EPUB Publications 3.0 http://www.idpf.org/epub/30/spec/epub30-publications.html
Spécification pour le conditionnement	EPUB Open Container Format (OCF) 3.0 http://www.idpf.org/epub/30/spec/epub30-ocf.html
Spécification pour le contenu	EPUB Content Documents 3.0 http://www.idpf.org/epub/30/spec/epub30-contentdocs.html
Spécification de la variante fixed layout	EPUB 3 Fixed-Layout Documents http://www.idpf.org/epub/fxl/
Accessibilité	http://www.idpf.org/accessibility/guidelines
Structuration sémantique	http://www.idpf.org/epub/vocab/structure
Structuration Daisy	http://www.daisy.org/z3986/structure/SG-DAISY3/index.html
DTD DTBook 2005-3	http://www.daisy.org/daisyniso-standard-dtd-and-css-files
Référentiel d'enrichissement des métadonnées	BnF-ADM-2013-002273-01
Référentiel de livraison de document numérique	BnF-ADM-2013-077351-01

Comment créer des identifiants ark
pour des tests

BnF-ADM-2007-013608-02

2. INTRODUCTION AU FORMAT EPUB

Le format ePub est un format standardisé de publication numérique, normalisé par l'IDPF (International Digital Publishing Forum) depuis 2006. La version 2.0.1 a été publiée en 2007 et la version 3.0 en 2011.

La version 2 se caractérise par les choix techniques suivants :

- format orienté « flot de texte » recomposable (une variante avec mise en page fixe apparaît dans ePub 3),
- basé sur les normes ou standards suivants : XHTML, CSS, Unicode, SVG,
- incorporation de polices de caractères,
- table de navigation,
- métadonnées bibliographiques (au format Dublin Core).

La version 3 se caractérise principalement par les choix techniques suivants :

- une variante avec mise en page fixe (*fixed layout*) est introduite,
- les contenus sont codés en HTML 5,
- extension des mécanismes de définition des métadonnées,
- support de MathML,
- support de l'accessibilité,
- extensions multimédia (ePub 3) : contenus audio et vidéo, synchronisation texte-son, interactivité, etc.



Pourquoi la version 3 du format ePub ?

Le format ePub 3 présente l'avantage de réunir les avantages d'un format de livre numérique ouvert, grand public et interopérable, tout en étant doté des mécanismes garantissant un meilleur accès aux contenus numériques pour les personnes souffrant d'une déficience visuelle légère ou moyenne comme pour les aveugles et malvoyants. Ces mécanismes d'accessibilité reposent sur une structuration logique des contenus numériques, afin de favoriser la navigation dans les documents.

3. PROCESSUS DE PRODUCTION EPUB

Il existe principalement deux types de processus de livres numériques au format ePub :

- réfection de documents déjà numérisés,
- production concomitante à un marché de numérisation.



Les règles et consignes développées dans le présent document s'appliquent quel que soit le mode de production.

3.1 Réfection de documents déjà numérisés

Lors de la production de livres au format ePub à partir de documents déjà présents dans la bibliothèque numérique de la BnF, le prestataire s'appuiera sur les éléments suivants, contenus dans le paquet des documents numériques mis à disposition par la BnF :

- le manifeste numérique (format refNum),
- les scans image,
- l'OCR (s'il est disponible, format ALTO),
- la table (table des matières ou index, si disponible, au format tdmNum).

3.1.1 OCR en qualité éditoriale

S'ils existaient dans le document numérique en réfection, les fichiers ALTO du paquet numérique du document sont mis à jour avec les contenus texte corrigés lors de la production de l'ePub. Dans le cas contraire, des fichiers ALTO sont ajoutés au document numérique en réfection.



Cette action vise à ce que le mode texte de Gallica profite de la montée en qualité du texte réalisée lors de la production des livres numériques au format ePub.

3.2 Production concomitante à un marché de numérisation

Lors de la production de livres au format ePub au sein d'une prestation de numérisation, le prestataire applique le processus de production de son choix.

3.2.1 OCR en qualité éditoriale

Le prestataire garantit que les fichiers ALTO du paquet numérique des documents produits profitent de la montée en qualité du texte réalisée lors de la production ePub.

4. CHOIX DES DOCUMENTS A TRAITER EN EPUB

La sélection des ouvrages à convertir en ePub est réalisée en prenant en compte plusieurs critères, exposés dans les sections suivantes :

- critères techniques,
- critères bibliographiques,
- critères intellectuels.

4.1 Critères techniques

4.1.1 Origine des documents

Les documents éligibles à une production au format ePub peuvent provenir des collections de la BnF ou de celles de ses partenaires.

Par ailleurs, ils peuvent avoir été l'objet d'une numérisation antérieure ou non (cf. section 4) :

- document numérisé et océrisé : selon le taux de qualité OCR du document, on envisagera d'utiliser le contenu OCR déjà disponible ou d'appliquer une nouvelle océrisation au document. Le gain attendu de cette réocérisation est une amélioration du taux de qualité OCR, ce taux étant lié aux performances des logiciels OCR, et donc inversement proportionnel à l'ancienneté des campagnes de numérisation.
- document numérisé : si la qualité de numérisation convient, le document sera océrisé.
- document non numérisé : la production d'un ePub interviendra après le processus de numérisation et d'océrisation du document.

4.1.2 Qualité OCR des documents

Les limitations de la numérisation de masse en matière de qualité du texte obtenu par océrisation sont connues. Elles sont dues à la nature même des documents (anciens, dégradés, etc.) et aux limitations intrinsèques des logiciels OCR. Ces limitations conduisent donc à des taux de qualité OCR variables, qui même lorsqu'ils sont qualifiés de « HQ » (taux à 99,9 %), ne garantissent pas un texte parfait à 99,9 % sur l'ensemble du document mais sur un sous-ensemble de celui-ci. En effet, certaines parties du document sont traitées en OCR brut (sans reprise manuelle du texte), car elles sont considérées comme trop coûteuses à traiter :

- des zones sont produites en OCR brut ou en image : texte masqué derrière un tampon, zone de l'image avec courbure, texte manuscrit, tableau, paragraphe peu lisible, etc.
- certaines catégories de texte sont produites en OCR brut, au niveau du mot ou du caractère : chiffres, ponctuation, diacritiques, mots inexploitablement déclassés par l'OCR.

Or les livres numériques issus de documents patrimoniaux seront lus hors du contexte de la bibliothèque numérique Gallica, par des lecteurs ayant un référentiel en matière de lecture numérique différent de celui des usagers de Gallica, c'est-à-dire avec une faible tolérance vis-à-vis des coquilles.

La sélection technique des documents éligibles à une production au format ePub doit donc s'appuyer sur deux actions :

- Choisir des documents ayant été commandés et traités en OCR HQ (ou les faire traiter en OCR HQ) et dont on sait ou suppose que ce taux HQ s'applique sur la quasi-totalité des contenus (c'est-à-dire des documents ayant peu de zones déqualifiées en OCR brut : tableaux, marges rognées, zones de courbure, etc.).
- Appliquer une montée en qualité du texte à ces documents, préalablement à la production ePub. Cette reprise s'applique notamment :
 - aux chiffres, ponctuations et diacritiques,
 - aux mots déqualifiés par l'OCR : les mots déqualifiés par l'OCR suivant le critère de *Word Confidence WC* < 0,5 sont à produire en haute qualité à condition que le texte soit lisible pour l'œil humain,
 - aux zones de texte déqualifiées par le processus OCR.

Cas des ePub « aménagés »

La qualité des fonds patrimoniaux demandés en OCR HQ ne permet pas toujours de monter la qualité du texte au niveau éditorial nécessaire à la conversion ePub. En accord avec la BNF, le prestataire peut proposer des aménagements pour améliorer le taux de convertibilité des documents vers le format ePub. Ces aménagements visent les documents HQ de bonne qualité « globale » présentant au maximum 5 % de pages impactées par des défauts (tâches, inclinaisons, migration d'encre, transparence, différence de contraste, texte tronqué).

L'aménagement consiste à différencier dans l'ePub les parties de texte perturbées, par exemple à l'aide d'une typographie particulière, et à apposer sur le texte un lien aller-retour vers une image de la partie concernée, l'image étant extraite du document origine. Les images seront par exemple positionnées à la fin de l'ePub.



Si pendant que l'air dans la nuit d'un jour de réjouissance, est parsemé de brillantes Fusées ou d'Etoiles artificielles que l'on jeteroit de dessus la Mer, nous entendions les Poissons dire en leur langage, Que de malheurs, que de desastres nous présagent ces Feux & ces Fusées ! n'aurions-nous pas raison de leur dire, Pauvres poissons, on ne pense pas à vous, vivez en repos, nos Fusées & nos Feux ne vous présagent aucun mal ? Appliquons-nous à nous-mêmes ce que nous leur dirions.

Je remarque que tous ceux qui par l'apparition des Comètes & des nouvelles Etoiles, ont prédit les révolutions des Etats, les changemens dans la Religion, ont esté de faux Prophetes. L'exemple du grand Tychobrahé suffira. Il avoit osé prédire un entier bouleversement de toute choses sur la Terre par l'apparition de la nouvelle Etoile de 1572. Cependant tout le monde sçait qu'elle a esté l'événement de la Prophetie de ce Prince des Astronomes.

Kepler, le plus célèbre des Astronomes Coperniciens, fut aussi grand menteur que Tychobrahé, lors qu'en parlant de la Comète de 1607, il juroit hardiment qu'elle avoit esté allumée entre les autres Astres, pour avertir les Hommes que Dieu avoit résolu de faire périr dans peu de temps, une grande partie du Genre-Humain.

Ceux qui prétendent que ces Etoiles chevelués sont toujours accompagnées de quelques grands malheurs, n'ont pour preuves que quelques inductions ; mais par une mesme forte de raisonnement, je pourrois conclure que les Comètes annoncent toujours quelque grand bonheur à la Terre.

Il avoit osé prédire un entier bouleversement de toute choses sur la Terre par l'apparition de la nouvelle Etoile de 1572. Cependant tout le monde sçait qu'elle a esté l'événement de la Prophetie de ce Prince des Astronomes.

Kepler, le plus célèbre des Astronomes Coperniciens, fut aussi grand menteur que Tychobrahé, lors qu'en parlant de la Comète de 1607, il juroit hardiment qu'elle avoit esté allumée entre les autres Astres, pour

Retour

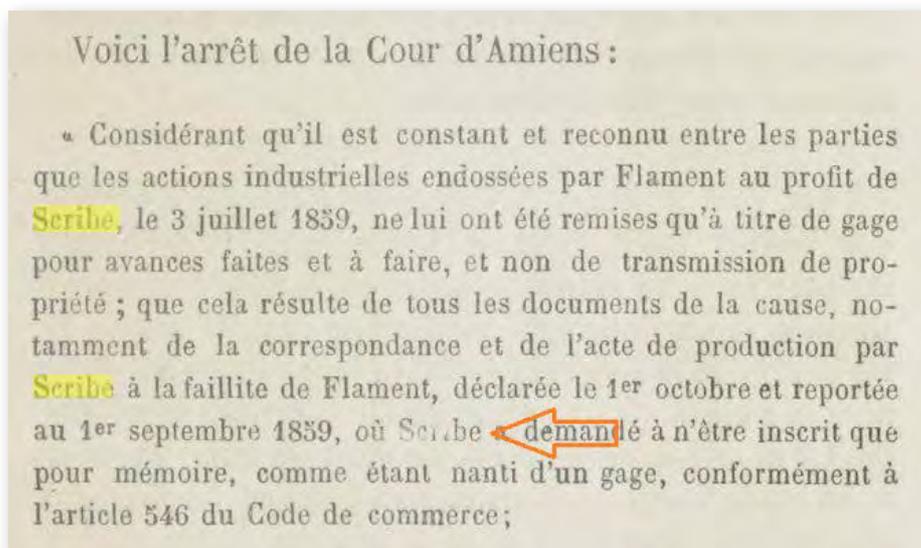
Principe de fonctionnement d'un ePub aménagé

Si la surface du(des) paragraphe(s) impacté(s) par le(les) défaut(s) représente moins de 50 % de la surface de la page, le lien enverra vers l'image d'origine du (des) paragraphe(s) concerné(s), sinon, le lien enverra vers l'image de la page complète.



ATTENTION

POUR ETRE TRAITEES SELON LE MECANISME DE L'Epub AMENAGE, LES PARTIES DE TEXTE PERTURBEES DOIVENT POSER UN REEL PROBLEME DE LISIBILITE (CF. SECTION 5.3.10). PAR EXEMPLE, LE CAS SUIVANT NE RELEVE PAS DE CE MECANISME, LE MOT AFFECTE PAR LE DEFAUT (« SCRIBE ») POUVANT ETRE AISEMENT DEDUIT DU CONTEXTE DU PARAGRAPHE.



Cas hors ePub aménagé

<http://gallica.bnf.fr/ark:/12148/bpt6k64658056/f63.image.r=scribe>

4.2 Critères bibliographiques

Les contraintes techniques propres au format ePub standard (flot de texte recomposable) imposent de privilégier les documents caractérisés par des contenus majoritairement textuels et une mise en page « simple » :

- pas de multicolonnage, d'imbrication texte/image complexe (habillage),
- pas ou peu de tableaux, de formules mathématiques, etc.

Les documents éligibles à une production au format ePub doivent donc être préférentiellement des monographies ou des périodiques de type revue, non presse (pas plus de deux colonnes de texte) :

- pas de monographies en plusieurs volumes,

- pas d'œuvres complètes ni de périodiques de type presse — leur traitement bibliographique n'est pas encore étudié.
- pas d'ouvrages scientifiques ou techniques riches en formules, tableaux, etc.

Par ailleurs, les documents ne doivent pas présenter une typographie complexe qui serait indispensable à la compréhension de l'œuvre. Ces caractéristiques typographiques sont en effet difficilement identifiées lors de la phase d'OCR.

Afin de faciliter cette étape de filtrage, la BnF et le prestataire pourront par exemple établir un catalogue des types de documents à rejeter pour la production ePub.

4.3 Critères intellectuels

L'objectif étant de produire des livres numériques, il convient de privilégier des contenus se prêtant bien à la lecture sur liseuses et tablettes numériques et aux utilisateurs potentiels de ces appareils. De manière schématique, les ouvrages de littérature, d'histoire, de vulgarisation scientifique, etc. seront préférés aux bibliographies, aux journaux officiels, etc. L'accès à ces derniers types documentaires est déjà permis via Gallica ; par ailleurs, la consultation web se prête bien mieux à un travail de recherche que le livre numérique, les liseuses et les logiciels de lecture de livre numérique étant des environnements informatiques plus « fermés » que les navigateurs web.

La valeur intellectuelle des contenus dédiés à une conversion ePub pourra être établie à l'aide de deux leviers :

- a priori, d'après le genre, la cote bibliographique, la thématique, etc. Ces critères peuvent être établis en début de marché, de tranches, voire au fil du marché ;
- document par document, sur demande spécifique des équipes de conservation.



LA PERTINENCE DE LA DIFFUSION D'UN DOCUMENT DONNE SOUS FORME DE LIVRE NUMERIQUE SE POSE EGALEMENT EN TERMES DE BUDGET, PUISQUE LA PRODUCTION AU FORMAT EPUB CAUSE UN SURCOUT IMPORTANT PAR RAPPORT A UNE NUMERISATION CLASSIQUE.

4.4 Recommandations générales

Les règles de qualification des documents au titre de la conversion ePub s'appuient donc sur des critères à la fois techniques (par ex. la qualité de la numérisation, le taux qualité OCR), bibliographiques (la nature des documents) et éditoriaux. L'application de ces règles est affaire de compromis et devrait conduire, idéalement, à un choix prenant en compte à la fois la valeur intellectuelle de l'œuvre et les difficultés techniques (tel celui, par exemple, de choisir de produire en EPUB un court recueil de poésie qui demanderait un effort de mise en page mais nécessiterait en contrepartie un faible volume de corrections).

4.4.1 Mode de sélection

Afin de limiter le manque de souplesse de tout marché de numérisation de masse, il est recommandé d'opérer des choix de documents s'appuyant sur les critères

énoncés ci-avant et aboutissant à la constitution de lots les plus homogènes possibles, sur lesquels le prestataire pourra appliquer un processus industriel.

L'élaboration d'un lot de documents éligibles à une production au format ePub pourra par exemple suivre les étapes suivantes :

- critères bibliographiques et/ou intellectuels :
 - choix d'une cote bibliographique (ouvrages du même genre),
 - filtrage selon la date de publication (par exemple postérieure à une année x ou un siècle x , afin d'éliminer les documents anciens dont on sait qu'ils auront un faible taux qualité OCR),
 - filtrage selon le catalogue des rejets ePub,
- critères techniques :
 - filtrage selon le taux OCR HQ (afin d'éliminer les documents non-HQ),
 - filtrage selon le taux de couverture OCR (proportion des mots considérés comme illisibles relativement à la population totale des mots), afin de limiter la proportion de textes illisibles qu'il faudrait reprendre manuellement.

4.4.2 Avertissement éditorial

Afin de prévenir le lecteur quant au fait que le livre numérique qu'il a en main peut présenter des coquilles et des défauts de forme, il convient d'insérer un avertissement éditorial dans l'ePub. Il explique la nature particulière des documents patrimoniaux qui sont à la source de l'ePub et les limitations liées au processus de production lui-même.

Cet avertissement est rédigé par la BnF et fourni au prestataire.

4.4.3 Contrôle de la sélection

Certains des critères de sélection présentés ci-avant peuvent être appliqués de manière automatique par un prestataire, lorsqu'il lui a été confié la responsabilité d'extraire un lot de documents éligibles à une production au format ePub d'un corpus plus vaste. Il conviendra dans ce cas de contrôler cette sélection. Les critères non techniques pourront également être mis en avant en orientant le prestataire vers tels ou tels groupes de documents, ou en faisant des choix à partir de propositions du prestataire.

Dans le cas où la BnF opère elle-même la sélection, la question du contrôle ne se pose pas.

5. CONVERSION VERS LE FORMAT EPUB

La conversion vers le format ePub implique des tâches :

- de montée en qualité des contenus texte,
- de structuration logique des contenus,
- de mise en forme.

5.1 Montée en qualité des contenus texte

Cette montée en qualité (cf. section 5.1.2) s'applique en particulier :

- aux zones de texte produites en OCR brut,
- aux mots déqualifiés (typés comme illisibles),
- à certaines catégories de texte produites en OCR brut, au niveau du mot ou du caractère : chiffres, ponctuation, diacritiques, espaces, mots inexploitablement déclassés par l'OCR.



L'objectif est de garantir une qualité finale de type « éditoriale » aux ePub produits.

5.1.1 Périmètre de la montée en qualité du texte

Cette montée en qualité ne concerne pas :

- les coquilles présentes dans le document original ;
- les textes (caractères, mots, zones) qui ne seraient pas lisibles dans le document original par un œil humain (cf. section 5.3.10, « Textes illisibles ») ;
- à certains contenus (notamment les tableaux complexes, les publicités, les formules de mathématiques ou de chimie), qui seront restitués en mode image (cf. sections 5.3.5 à 5.3.8) ;
- aux portions de texte composées dans un alphabet non latin, qui seront restituées en mode image (cf. section 5.3.10, « Textes en langue étrangère ») ;
- aux textes manuscrits, qui seront restitués en mode image (cf. section 5.3.10, « Textes en langue étrangère »).



LES ERREURS D'OCR AYANT POUR ORIGINE UN DEFAUT DE QUALITE DU DOCUMENT PHYSIQUE (DANS LE CADRE D'UNE PRESTATION DE NUMERISATION) OU DE SON IMAGE NUMERIQUE (DANS LE CADRE D'UNE PRESTATION DE REFECTION) FONT PARTIE DU PERIMETRE DE LA MONTEE EN QUALITE ET DOIVENT ETRE CORRIGEEES PAR LE PRESTATAIRE. CES ERREURS SONT GENERALEMENT LIEES A DES PROBLEMES DE BRUIT DANS L'IMAGE, DE COURBURES, DE DEFAUT D'ENCRAGE, ETC.

MAIS LES TEXTES MANQUANTS OU ILLISIBLES DANS L'ORIGINAL NE FONT PAS PARTIE DU PERIMETRE DE LA MONTEE EN QUALITE.

souviens encore que lorsque les larges pieds des vignons foulaient la vendange au château de mon père, fuyais épouvantée de la senteur des cuves, et que j'allais bien loin m'asseoir sur quelque hauteur pour respirer le vent du ciel.

— Avec vos cheveux que le soleil empourpre et dans ce moment vous eussiez pourtant fait une fort belle Érigone, reprit-il galamment. Croyez-moi, votre dédain pour le breuvage que tous les peuples ont appelé divin, a quelque chose d'affecté et de manière qui n'est pas digne de vous.

— Mais je n'affecte rien, je vous jure; c'est en moi un instinct de répulsion, et le jour où cette répugnance cesserait, je vous promets d'essayer de boire avec vous.

— Oh! reprit-il, quelle bonne femme vous êtes! N'est-ce pas, vous ne croirez pas ce qu'on vous dira de moi: que je m'abrutis, que je me jette tête baissée dans cet oubli de l'ivresse? Non, non, je vois sciemment ce que je fais et ce que je veux quand parfois je m'abandonne. Chère marquise, si jamais votre cœur se déchiré, ne regardez pas un homme du peuple ivre, chantant et riant dans sa misère, cela vous donnerait le vertige et l'envie de l'imiter.

— C'est un expédient aveugle et matériel, lui dis-je; ne peut-on s'étourdir par l'amour, par le dévouement, par le patriotisme, par la gloire?

— J'ai essayé de tout, et l'oubli seules là, répliquait-il en frappant la bouteille du revers de ses doigts

Texte hors périmètre qualité dans une image en réfection

<http://gallica.bnf.fr/ark:/12148/bpt6k6124268h/f53.image>

Cette montée en qualité est évaluée à l'aide de deux taux qualité :

- taux des erreurs majeures résiduelles,
- taux des erreurs mineures résiduelles.

Ces taux qualité sont calculés :

- au mot, soit par exemple un mot erroné tous les 10 000 caractères pour un taux de 99,99 % ;
- sur les seuls mots présents dans le périmètre de la montée en qualité.

5.1.2 Erreurs majeures

Un mot est considéré comme en erreur majeure dans les cas suivants :

- un(des) caractère(s) manquant par rapport au texte du document original : « erreur » pour « erreur » ;
- un(des) caractère(s) inséré(s) à tort car absent(s) du texte du document original : « erreurs » pour « erreur » ;
- un(des) caractère(s) différent(s) du texte du document original : « erreun » pour « erreur » ;
- deux mots distincts dans le texte du document original sont collés ;
- un mot du texte du document original est séparé en plusieurs composantes.

Ce taux est calculé à l'aide de la formule suivante :

$$\frac{\text{cardinal des mots erronés parmi les mots dans le périmètre de la qualité garantie}}{\text{cardinal des mots dans le périmètre de la qualité garantie}}$$



Les caractères à prendre en compte pour le calcul de ce taux sont les suivants :

- lettres [a-z]
- chiffres [0-9]

5.1.3 Erreurs mineures

Un mot ou un caractère est considéré comme en erreur mineure dans les cas suivants :

- erreurs sur les signes de ponctuation : une ponctuation est erronée si elle est manquante, insérée à tort ou différente de l'originale
- caractères non alphanumériques : un caractère non alphanumérique est erroné s'il est manquant, inséré à tort ou différent
- erreurs d'espace : espace doublé, espace insécable manquant avant une ponctuation, etc.
- erreurs de casse : *a* pour *A*
- erreurs d'accents : *é* pour *è*



Les caractères à prendre en compte pour le calcul de ce taux sont les suivants :

- toutes les diacritiques (é, è, ê, etc.) des lettres [a-z]
- les signes de ponctuation : , . ; : ? ! " " « »
- les caractères non alphanumériques : # \$ £ % & ¥ @ () [] { } * + ± = / ÷ - - < > ° ~ • © § ¶
- les caractères espace : seules l'espace simple et l'espace insécable sont considérées.
- les capitales des lettres [a-z] (les petites capitales seront considérées comme des capitales).

Ce taux est calculé à l'aide de la formule suivante :

$$\frac{\text{cardinal des mots ou caractères erronés parmi les mots dans le périmètre de la qualité garantie}}{\text{cardinal des mots dans le périmètre de la qualité garantie}}$$

5.1.4 Valeurs des taux qualité

La montée en qualité du texte est évaluée à l'aide de deux taux qualité :

- taux des erreurs majeures,
- taux des erreurs mineures.

Des taux spécifiques s'appliquent à certains documents pour lesquels le prestataire pourra demander l'application d'un taux qualité inférieur au taux « qualité éditoriale » :

- critères bibliographiques : typographie ancienne, ancien français, dialecte ou patois français, présence importante de textes dans des langues autres que le français ;
- critères physiques quand le prestataire réalise la numérisation (ex : courbure inévitable, cf. section 4.1.2) ;
- problèmes liés à la qualité des images scannées lorsqu'il s'agit d'une prestation de réfection.

Les valeurs de ces taux qualité sont définies dans le tableau suivant :

Typologie	Taux attendu (mesuré au mot)
Qualité éditoriale « argent »	Erreurs majeures : < 0,06 % (taux qualité : 99,94 %)
	Erreurs mineures : < 0,1 % (taux qualité : 99,9 %)
Qualité éditoriale « bronze »	Erreurs majeures : < 0,1 % (taux qualité : 99,9 %)
	Erreurs mineures : < 0,17 % (taux qualité : 99,83 %)

5.2 Structuration

La structuration et le typage des entités textuelles étant plus riches pour l'ePub que pour l'ALTO, des spécifications adaptées pour la production des livres numériques sont à déployer par le prestataire. Cette montée en qualité concerne en particulier la structuration logique des contenus et donc leur mise en forme (la mise en forme s'appuyant sur les informations de structuration).

5.2.1 Périmètre de la montée en qualité

Cette montée en qualité concerne la totalité du contenu excepté :

- les contenus (notamment les tableaux complexes, les publicités, les formules de mathématiques ou de chimie), qui seront restitués en mode image (cf. sections 5.3.5 à 5.3.8),

La montée en qualité est évaluée par contrôle automatique et par contrôle visuel (cf. section 9).

5.2.2 Segmentation

La montée en qualité de la segmentation concerne principalement :

- le respect de l'ordre de lecture du document d'origine : l'ordre de lecture produit par l'OCR peut s'avérer faux et le bon ordre de lecture doit être rétabli,
- la correction de la segmentation du texte : sauts de ligne, sauts de paragraphe, césure scindée entre deux blocs de texte ou deux pages, etc.

5.2.3 Structuration logique

La structuration logique consiste à qualifier sémantiquement les différents contenus du document numérique :

- identification de la langue des contenus ;
- identification des types de page du document (pages de titre et faux-titre, débuts de chapitre, index, etc.). Ce typage pourra être repris des informations contenues dans le paquet du document numérique (manifeste RefNum) ;



Les pages de logo (type L dans le manifeste RefNum) sont à supprimer du contenu de l'ePub.

- identification des numéros de page
- distinction des différents niveaux de structuration du document numérique : partie, chapitre, titre, etc. Soit par exemple :
 - partie : titre de niveau 1 ; chapitre : titre de niveau 2 ; section : titre de niveau 3
 - chapitre : titre de niveau 1 ; section : titre de niveau 2
 - article : titre de niveau 1 ; rubrique : titre de niveau 2
 - ...
- distinction des types de paragraphe de tous les contenus textuels : texte courant, remarque, encadré, note de bas de page, légendes de figure, listes numérotées et tout autre élément doté d'une numérotation séquentielle dans le document numérique, etc.
- distinction des types de contenus en mode image : figures, illustrations, traits de séparation, ornements, etc. Ce typage pourra être repris des informations contenues dans les fichiers ALTO.



Les mécanismes d'accessibilité d'ePub 3 reposent sur une structuration des contenus numériques. Cela implique que les producteurs de contenus numériques publient des contenus *structurés et décrits*.

Dans ePub 3, cette structuration logique s'incarne à l'aide :

- d'un balisage HTML 5 de type sémantique : <section>, <article>, <aside>, <figure>, etc.
- de l'attribut d'inflexion sémantique epub:type, permettant de préciser la nature des diverses composantes de contenu : frontmatter, bodymatter, backmatter, chapter, toc, glossary, index, footnote, prologue, epilogue, preface, appendix, etc.

Les sections suivantes présentent le détail des tâches de structuration et d'identification à réaliser.

Identification de la langue

La langue principale de l'œuvre est connue et elle est obligatoirement renseignée dans les métadonnées de l'ePub (cf. section 6.1.7).

Mais les portions de contenu hors langue principale peuvent être détectées par les moteurs OCR. Cette information devra être reprise dans les contenus HTML ePub grâce à l'attribut `xml:lang`, au niveau du paragraphe comme du mot.

Cette information sera également renseignée dans le flux ALTO grâce à l'attribut `language`.

Macro-structuration logique

L'identification des grandes composantes du document (couverture, liminaires, annexes, corps de l'ouvrage, etc.) est nécessaire à la production d'ePub 3 accessibles. Les composantes à identifier sont définies dans le tableau suivant :

Structuration macro	A réaliser	Exemple HTML 5
Couverture	✓	<code><body epub:type="cover"></code> ... <code></body></code>
Liminaires (préface, prologue, avertissement, etc.)	✓	<code><body epub:type="frontmatter"></code> ... <code></body></code>
Contenu principal	✓	<code><body epub:type="bodymatter"></code> ... <code></body></code>
Fin d'ouvrage (annexes, index, etc.)	✓	<code><body epub:type="backmatter"></code> ... <code></body></code>

Structuration logique intermédiaire

La structuration logique de niveau intermédiaire vise à structurer logiquement le contenu. Elle permet de distinguer les différents niveaux de structuration du document numérique : partie, chapitre, titre, etc., et donc d'en faciliter la consultation.

Structuration intermédiaire	A réaliser	Exemple HTML 5
Volume, tome	Ce niveau n'existe pas dans les documents de la BnF, il est du ressort du catalogue.	
Partie	✓	<code><section epub:type="part"></code> <code><h1>Partie I</h1></code> ... <code><section epub:type="chapter"></code> <code><h2>Chapitre 1</h2></code> ... <code></section></code>

		</section>
Chapitre	✓	<pre><body> <section epub:type="chapter"> <h2>DISCOURS SUR LA NATURE DES POISSONS.</h2> ... </section> </body></pre>
Sous-chapitre	Ce niveau de structuration peut ne pas être utilisé	<pre><section epub:type="subchapter"> <h3>Titre 1.1</h3> <p classe="texte">Blabla</p> ... </section></pre>
Division	Ce niveau de structuration peut ne pas être utilisé	
Sections	✓	<pre><section> <h4>Titre 1.1.1</h4> <p classe="texte">Blabla</p> ... </section></pre>



Différence entre structuration orientée « rendu graphique » et structuration logique

Si l'on prend l'exemple de la structuration des niveaux de titre d'un document, une structuration orientée « rendu graphique » consiste simplement à marquer les niveaux de titre par des styles CSS et non à englober dans un balisage hiérarchique les portions de contenu concernées par les niveaux de titre.

Structuration orientée « rendu graphique » (ePub 2)

```
<p classe="chapitre">Chapitre 1</p>
<p classe="titre 1">Titre 1.1</p>
<p classe="texte">Blabla</p>
<p classe="texte">Blabla</p>
```

...

Structuration logique (ePub 3)

```
<section epub:type="chapter">
  <h1>Chapitre 1</h1>
  <section>
```

```

    <h2>Titre 1.1</h2>
    <p classe="texte">Blabla</p>
    <p classe="texte">Blabla</p>
    ...
  </section>
</section>

```

Structuration sémantique fine

La structuration sémantique fine permet de qualifier la nature des contenus. Le périmètre de cette tâche est donné dans le tableau suivant.

Structuration sémantique fine	A réaliser	Exemple HTML 5
Page de titre	✓	<pre> <section epub:type="titlepage"> <h1 epub:type="fulltitle">Titre</h1> ... </section> </pre>
Epigraphe, dédicace, avertissement, copyright, remerciements, crédits		
Préface, prologue, introduction, préambule, avant-propos		
Liste des illustrations	✓	<pre> <section epub:type="lot"> ... </section> </pre>
Liste des tables	✓	<pre> <section epub:type="loi"> ... </section> </pre>
Table des matières	✓	<pre> <section epub:type="toc"> ... </section> </pre>
Conclusion, épilogue		
Index	✓	<pre> <section epub:type="index"> ... </section> </pre>
Glossaire	✓	<pre> <section epub:type="glossary"> ... </section> </pre>
Bibliographie	✓	<pre> <section epub:type="bibliography"> ... </section> </pre>
Colophon, errata		



L'exemplarisation des ouvrages, qui permet d'identifier certaines catégories de pages dans le manifeste numérique du document (refNum), doit bien sûr être cohérente avec cette structuration pour les catégories de page suivantes : couverture, titre, table des matières, index.

Structuration micro

Les composantes de contenu de granularité élémentaire (titres, légendes, listes, etc.) doivent également être caractérisées.

Structuration micro	A réaliser	Exemple HTML 5
Paragrophes	✓	<p> ... </p>
Séparation forte entre deux paragraphes	✓	<hr class="transition"/>
Figures et légendes de figure	✓	<figure> <figcaption>Figure 1.1 ...</figcaption> </figure>
Notes de bas de page	✓	<p>...<a epub:type="noteref" href="#ndbp1">1 ... </p> <aside epub:type="footnote" id="ndbp1"> Texte de la note... </aside>
Notes marginales	✓	<aside epub:type="marginalia"> Texte de la note... </aside>
Encadrés	✓	<aside> ... </aside>
Listes à puces	✓	
Listes numérotées	✓	
Formules de maths/chimie	mode image	
Tableaux	✓	<table> <caption>...</caption>

		<thead> <tr> ...
Gras et italique	✓	 et <i>
Publicités, catalogues d'éditeur	mode image	

Structuration physique

Les numéros de page du document physique doivent être caractérisés et insérés dans le contenu numérique afin de permettre de générer la table des numéros de page (cf. section 5.5.3).

Structuration physique	A réaliser	Exemple HTML 5
n° de page (folio) marqué dans le contenu	✓	<p>...36 1...</p>

5.3 Mise en forme

La mise en forme des contenus s'appuie :

- sur la structuration logique des contenus,
- sur une ou plusieurs feuilles de styles de mise en page (au format CSS) proposées par le prestataire et validées par la BnF,
- sur les consignes énumérées dans les sections suivantes.

La BnF et le prestataire pourront évaluer, durant la phase de test, la possibilité d'adapter la charte graphique générique aux typologies documentaires particulières rencontrées.



LES SOUHAITS EXPRIMES CI-APRES SERONT ENVISAGES PAR LE PRESTATAIRE DANS UN SOUCI D'UNIVERSALITE DE LA SOLUTION, QUI DEVRA ETRE FONCTIONNELLE SUR LA MAJORITE DU PARC DES LISEUSES. PAR AILLEURS, SUR LES LISEUSES EN ECHEC QUANT A LA MISE EN FORME SOUHAITEE DU CONTENU, LE CONTENU LUI-MEME NE DEVRA PAS ETRE AFFECTE.

5.3.1 Styles de paragraphe

De manière générale, la typographie de l'ouvrage original ne pourra être restituée dans les ePub produits. Par exemple, l'indentation ne sera pas forcément reproduite. Le prestataire exposera donc sa capacité à restituer les différents alignements du texte :

- fer à gauche
- fer à droite
- centré
- justifié

- valeurs de marge
- etc.

Styles des titres

Les limitations de mise en forme typographique exposées ci-avant ne s'appliquent pas aux titres de section présents dans la table des matières de l'ePub, qui devront être mis en page avec un rendu graphique adapté, en fonction de leur place dans la hiérarchie des niveaux de titre.

Le prestataire proposera des solutions visant à empêcher :

- que la césure s'applique aux titres (de chapitre ou de section),
- qu'un titre soit séparé du paragraphe suivant (et donc affiché seul en bas de la fenêtre de lecture).

Les césures dans les titres sont à éviter, sauf dans les cas où la césure est nécessaire pour restituer la mise en page du document originel. Dans l'exemple suivant, un retour à la ligne forcé doit être inséré dans le titre de chapitre.



JAN MARTIN SECRETAIRE DE MONSEI-
gneur le Reverendissime Cardinal de Lenoncourt.
AUX LECTEURS.



I vous desirez, Messeigneurs, entendre à peu de parolles ce qui est contenu en cest œuvre, sachez que Poliphile dict avoir veu en songe des choses admirables, entre lesquelles il en décrit plusieurs antiques dignes de memoire, comme Pyramides, Obelisques, grandes ruines d'edifices, la difference des colonnes, leurs mesures, piedestalz, bazes, et chapiteaux dont elles sont ornées. Puis les architraves, frizes, cornices, et frontispices avec leurs ouvrages. Un grand cheval, un Elephant de merveilleuse grandeur, un Colosse, et une porte magnifique, avec son plant, ordonnance, moulures, et besongne de taille. Apres comment cinq belles Nymphes le menerent



JAN MARTIN SECRETAIRE DE MONSEI-
gneur le Reverendissime Cardinal de Lenoncourt.
AUX LECTEURS.



I vous desirez, Messeigneurs, entendre à peu de parolles ce qui est contenu en cest œuvre, sachez que Poliphile dict avoir veu en songe des choses admirables, entre lesquelles il en décrit plusieurs antiques dignes de mémoire, comme Pyramides, Obélisques, grandes ruines d'edifices, la différence des colonnes, leurs mesures, piedestalz, bazes, et chapiteaux dont elles sont ornées. Puis les architraves, frizes, cornices, et frontispices avec leurs ouvrages.

...

Styles de liste

Les listes à puces de l'original devront être restituées.

Les listes numérotées de l'original devront également être restituées avec la bonne numérotation.

Genres documentaires

Les styles documentaires suivants devront être restitués dans leurs principales caractéristiques de mise en page et de typographie : poésie, théâtre. Il s'agit essentiellement de préserver dans le rendu graphique de l'ePub :

- les marges et les alignements de paragraphe,
- les espaces interparagaphes,
- le corps de la police des textes, ses enrichissements (gras, italique),
- les signes typographiques spécifiques (par exemple le tiret de dialogue).



NOTES

On ne cherchera pas à reproduire la mise en page du document d'origine dans toutes ses caractéristiques, mais simplement celles qui sont essentielles au genre documentaire (par exemple des paragraphes au fer à gauche pour les vers d'un poème).



EXEMPLE

FABLES DE LA FONTAINE ¹



I. — LA CIGALE ET LA FOURMI.

La cigale, ayant chanté
Tout l'été,
Se trouva fort dépourvue ²
Quand la bise ³ fut venue :
Pas un seul petit morceau ⁴
De mouche ou de vermisseau ⁵.
Elle alla crier famine
Chez la fourmi sa voisine,
La priant de lui prêter
Quelque grain ⁶ pour subsister
Jusqu'à la saison nouvelle ⁷.
« Je vous paierai, lui dit-elle,



FABLES DE LA FONTAINE¹

I. – LA CIGALE ET LA FOURMI.

La cigale, ayant chanté
Tout l'été,
Se trouva fort dépourvue²
Quand la bise³ fut venue :
Pas un seul petit morceau⁴
De mouche ou de vermisseau⁵.
Elle alla crier famine
Chez la fourmi sa voisine,
La priant de lui prêter
Quelque grain⁶ pour subsister
Jusqu'à la saison nouvelle⁷.
« Je vous paierai, lui dit-elle,

...

Exemple de mise en page d'un poème

<http://gallica.bnf.fr/ark:/12148/bpt6k5449454s/f7.image>



ACTE PREMIER.

Le théâtre représente une place publique. A droite, au premier plan, une grille de jardin, attenant à une maison dont la porte d'entrée est près de la grille. Un balcon en saillie est au premier étage, entre la porte et le tournant de la maison ; une jalousie et un rideau masquent la fenêtre. A gauche, au premier plan, une auberge avec une tonnelle au-devant, sous laquelle se trouvent une table et un banc.

SCÈNE PREMIÈRE.

CÉLIO, PIPPO.

(Ils entrent par la droite, du plan au-dessus de la maison.)

CÉLIO.

Eh bien, Pippo, tu viens de voir Marianne ?

PIPPO.

Oui, monsieur.

CÉLIO.

Que t'a-t-elle dit ?



ACTE PREMIER.

Le théâtre représente une place publique. A droite, au premier plan, une grille de jardin, attenant à une maison dont la porte d'entrée est près de la grille. Un balcon en saillie est au premier étage, entre la porte et le tournant de la maison ; une jalousie et un rideau masquent la fenêtre. A gauche, au premier plan, une auberge avec une tonnelle au-devant, sous laquelle se trouvent une table et un banc.

SCÈNE PREMIÈRE.

CÉLIO, PIPPO.

(Ils entrent par la droite, du plan au-dessus de la maison.)

CÉLIO.

Eh bien, Pippo, tu viens de voir Marianne ?

PIPPO.

Oui, monsieur.

...

Exemple de mise en page d'une pièce de théâtre

<http://gallica.bnf.fr/ark:/12148/bpt6k166654c/f7.image>

5.3.2 Styles de caractère et typographie

La montée en qualité des contenus texte concerne aussi les enrichissements typographiques du texte : application de polices italique ou grasse, notations exposant ou indice, insertion de caractères spéciaux, etc.

Enrichissements typographiques (option)

Les enrichissements typographiques du texte (gras, italique, souligné) seront restitués.

Indice et exposant

Les indices et les exposants seront produits conformément à l'original.

Police et taille

La police et la taille originales du texte ne sont pas reproduites. Ces informations de style seront fournies par une feuille de style numérique (au format CSS) proposée par le prestataire et validée par la BnF.

Capitales et petites capitales

Les capitales sont reproduites. Les petites capitales sont reproduites sous forme de capitales.

Espaces

Les différentes espaces typographiques ne sont pas reproduites (espace fine, cadratin, demi-cadratin, etc.).

Cependant, lorsque dans le document d'origine, une espace est présente entre un mot et un signe de ponctuation, une espace insécable sera utilisée afin d'éviter, dans l'ePub produit, que le signe de ponctuation soit rejeté à la ligne.

Caractères spéciaux

De manière préférentielle, les caractères spéciaux à restituer le seront sous la forme de caractères pris dans une ou plusieurs polices incluses dans le document et spécifiés grâce à leur codage Unicode. Si nécessaire, un symbole (par exemple une puce dotée d'un dessin particulier) pourra être remplacé par un équivalent.

Les seuls caractères spéciaux restitués en mode image seront :

- les lettres ornées (voir section 5.3.4) ;
- les caractères ou symboles difficiles à identifier et donc à reproduire sous forme de point de code Unicode ou de caractère pris dans une police incorporée.

Signes typographiques

Les tirets d'incise, d'énumération et de liste (en général un tiret demi-cadratin) sont à produire dans le flot de texte avec le caractère Unicode –

Les tirets de dialogue, les tirets longs (en général un tiret cadratin) sont à produire dans le flot de texte avec le caractère Unicode —

Les puces de forme standard (•) sont à produire dans le flot de texte avec le code •

Les guillemets présents dans le document d'origine devront être reproduits dans l'ePub, sauf cas des guillemets répétés (cf. section 5.3.10).

Les règles typographiques françaises modernes devront être respectées, avec notamment l'insertion d'une espace insécable avant les signes ; : ! ? » et après le signe «.

Césure

Le prestataire évaluera la capacité de la norme ePub 3 à composer des textes respectant les règles de césure de la langue française. Après concertation avec la BnF, les mécanismes disponibles seront mis en œuvre ou abandonnés, selon la qualité de leur rendu typographique.

5.3.3 Illustrations

Les illustrations du document original sont conservées dans l'ePub.



Dans les fichiers ALTO, les illustrations sont des éléments Illustration.

Concernant le mode colorimétrique des images restituées dans l'ePub, les règles suivantes sont appliquées :

Images dans le document original	Images dans l'ePub	Commentaires
couleurs	couleurs	
niveaux de gris	niveaux de gris	
noir et blanc	niveaux de gris	Les niveaux de gris sont obtenus lors du redimensionnement de l'image.



Les images seront produites en respectant ces critères :

- L'orientation des images restera conforme à l'original.
- Une image pleine page dans l'original sera mise en page dans l'ePub avec un saut de page avant l'image et un saut de page après.
- La taille relative de l'image affichée dans l'ePub par rapport à celle de la page originale sera prise en compte afin de conserver le même ratio dans la mise en forme ePub.
- La taille effective de l'image (sa définition, en pixels) sera de 1,5 fois la taille de l'image affichée, afin de permettre à l'utilisateur de zoomer (avec les liseuses et logiciels de lecture ePub qui bénéficient de cette fonctionnalité).

Lorsqu'une illustration est incluse dans un paragraphe dans le document original, elle doit être insérée *après* le paragraphe dans l'ePub. La figure suivante montre un tel cas.

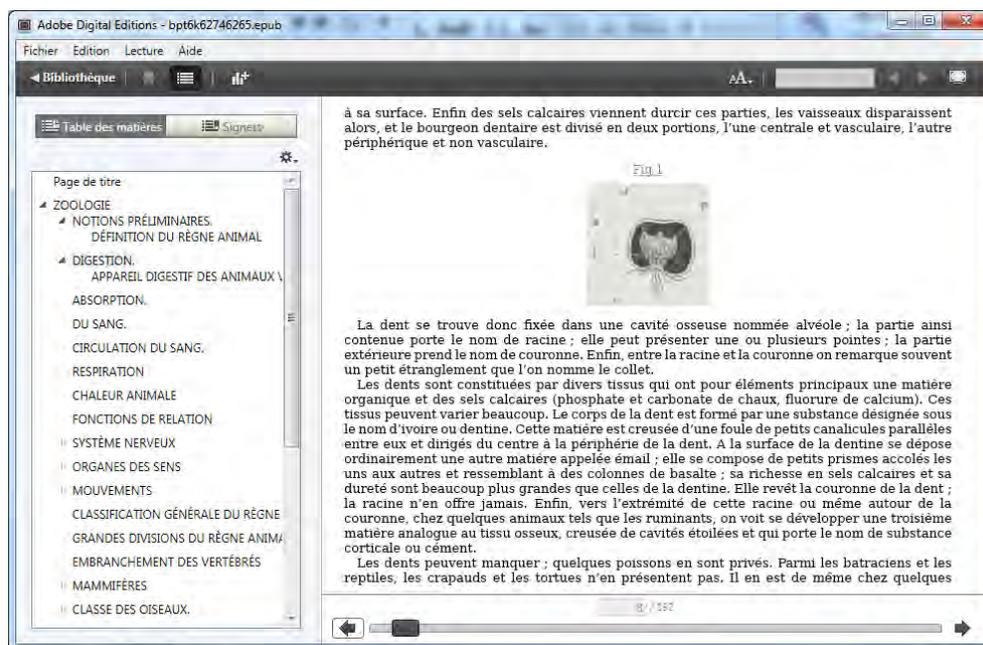


<http://gallica.bnf.fr/ark:/12148/btv1b8600206q/f38.zoom>

Légendes

Les légendes des illustrations, lorsqu'elles existent (typées CAPTION dans l'ALTO), sont placées avant l'illustration.

EXEMPLE



Placement des légendes de figure

OBLIGATOIRE

Chaque légende doit être groupée avec son illustration, à l'aide d'un codage XHTML et CSS adéquat, afin d'éviter que l'une ou l'autre ne soit séparée lors de la lecture, selon les configurations de mise en page.

Cas des illustrations habillées

Le format ePub permettant le redimensionnement dynamique du texte, il est difficile de restituer des illustrations habillées par du texte.

NOTES

Ce constat sera réévalué par le prestataire en fonction des capacités de mise en page de la version 3 du format ePub. Après concertation avec la BnF, un choix sera fait quant au traitement des illustrations habillées. La suite de cette section traite du cas où les illustrations ne seraient pas habillées.

Une illustration habillée par du texte correspond généralement à un codage ALTO particulier : illustration capturée en ComposedBlock avec des TextBlock pour le texte et un Illustration pour l'illustration :

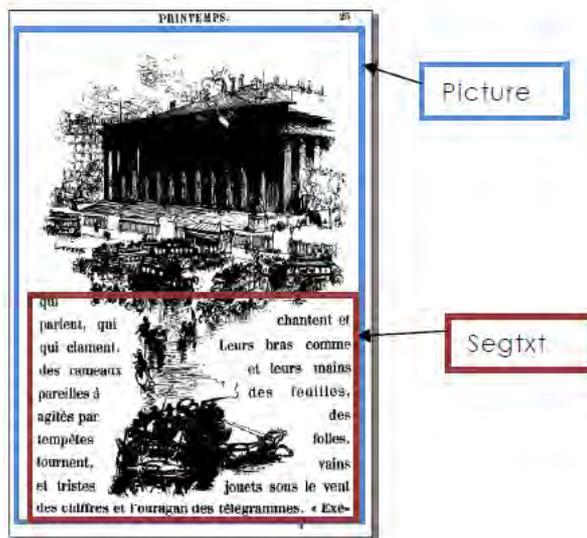


Image habillée par du texte

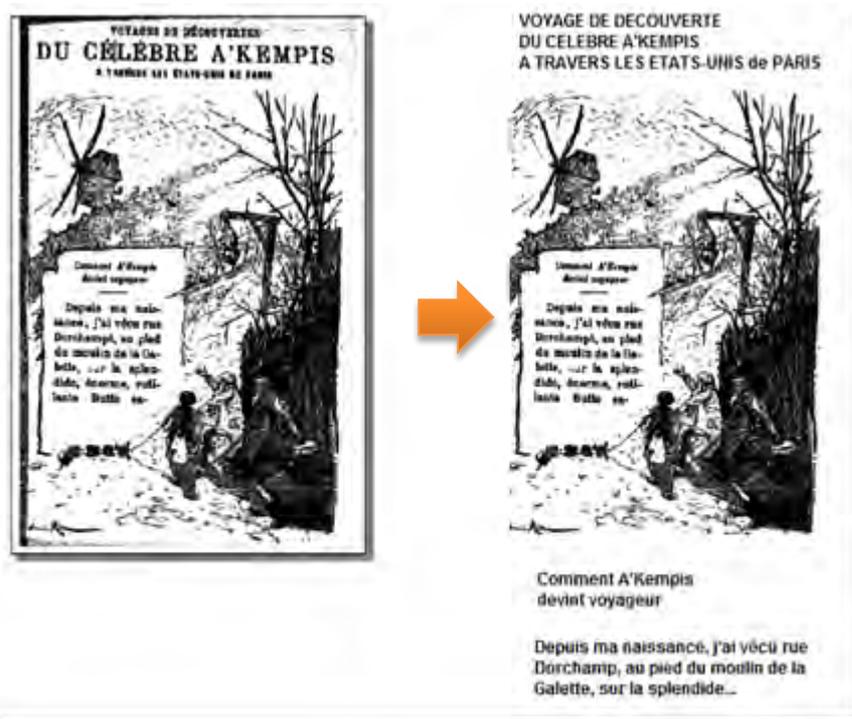
La règle de production est la suivante : dans l'ePub, le texte sera placé après l'illustration. L'illustration sera soit détournée, soit traitée de façon à supprimer le texte.



Ce traitement ne s'applique pas aux lettrines (cf. section 5.3.4), pour lesquelles des règles spécifiques s'appliquent.

Dans le cas où le texte est entouré par l'illustration, le même principe doit être appliqué, mais le texte n'a pas à être supprimé de l'illustration.

EXEMPLE



Texte inclus dans une image

5.3.4 Eléments graphiques

Les éléments graphiques tels que culs-de-lampe, tampons, traits de séparation ornés, décorations, etc., ne sont *pas* conservés dans l'ePub.

NOTES

Dans les fichiers ALTO, les éléments graphiques sont des `GraphicalElement`.

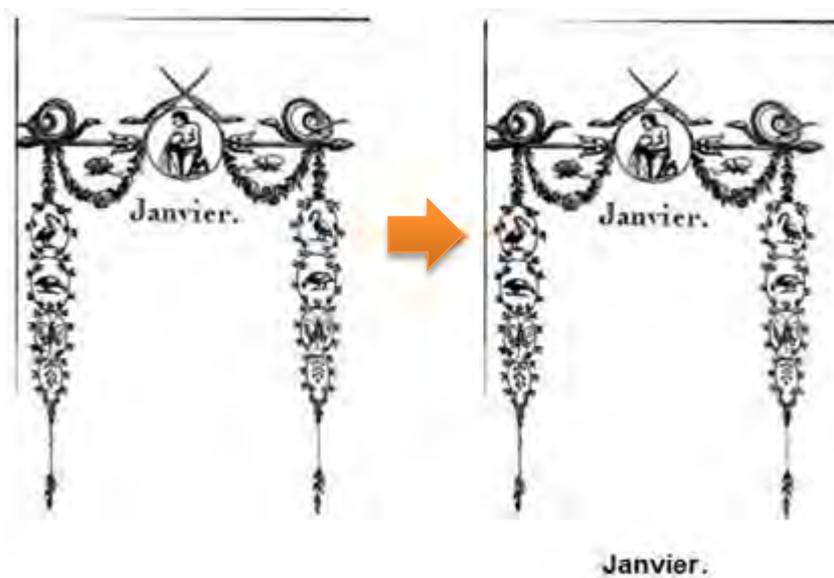
Il existe deux exceptions à cette règle :

- les décorations incluant du texte et les décorations de tête de chapitre devront être conservées (sans suppression du texte) et le texte inclus sera placé après la décoration.

Ce cas de figure correspond généralement à un codage ALTO particulier : décoration capturée en `ComposedBlock` avec un `TextBlock` pour le texte et un `GraphicalElement` autour des ornements :



 **EXEMPLE**



Exemple de décoration avec texte inclus



Exemple de décoration de tête de chapitre

<http://gallica.bnf.fr/ark:/12148/btv1b8600236v/f23.image>

 **NOTES**

Comme pour les illustrations, les capacités de mise en page de la version 3 du format ePub orienteront le traitement des décorations habillées par du texte.

- les traits de séparation et les culs-de-lampe placés entre deux paragraphes doivent être remplacés par un saut de ligne (cf. « Structuration micro », section 5.2.2).

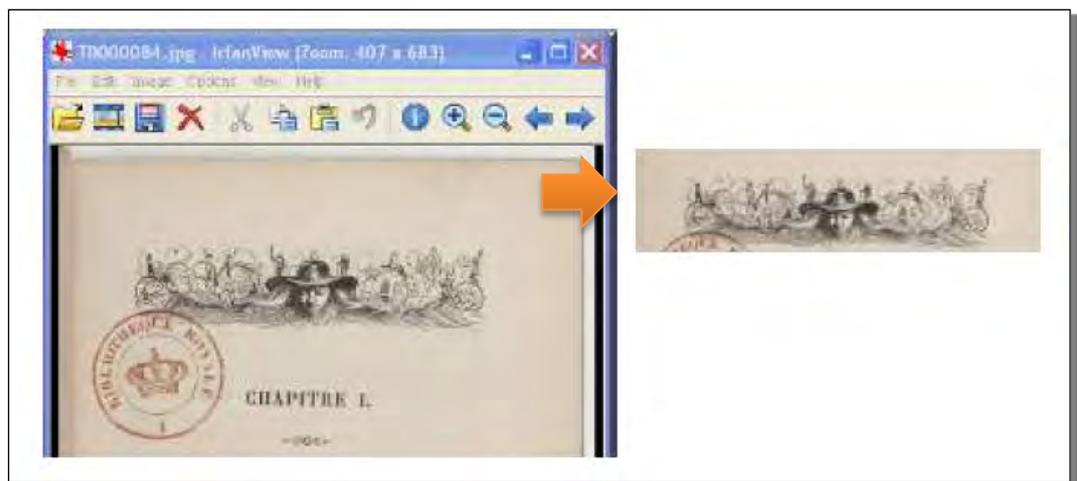


Dans les fichiers ALTO, ce cas est décrit avec un élément GraphicalElement doté de l'attribut TYPE="transition".

Tampons

Les tampons apposés sur l'original par la bibliothèque doivent être supprimés. Lorsqu'ils couvrent du texte, le texte masqué par le tampon doit être ressaisi. Par contre, si le texte derrière le tampon est illisible, la zone sera prise en mode image.

Lorsqu'un tampon touche une illustration, on segmente l'illustration sans tenir compte du tampon :



Lettrines

Les lettres ornées devront être conservées.

Les lettrines simples de petite taille seront ancrées au début du paragraphe qu'elles introduisent et le texte habillera la lettrine.



ON père avait un petit bien situé dans la province de Nottingham. J'étais le troisième de ses cinq fils. Il m'envoya au collège d'Emmanuel, à Cambridge, à l'âge de quatorze ans. ...



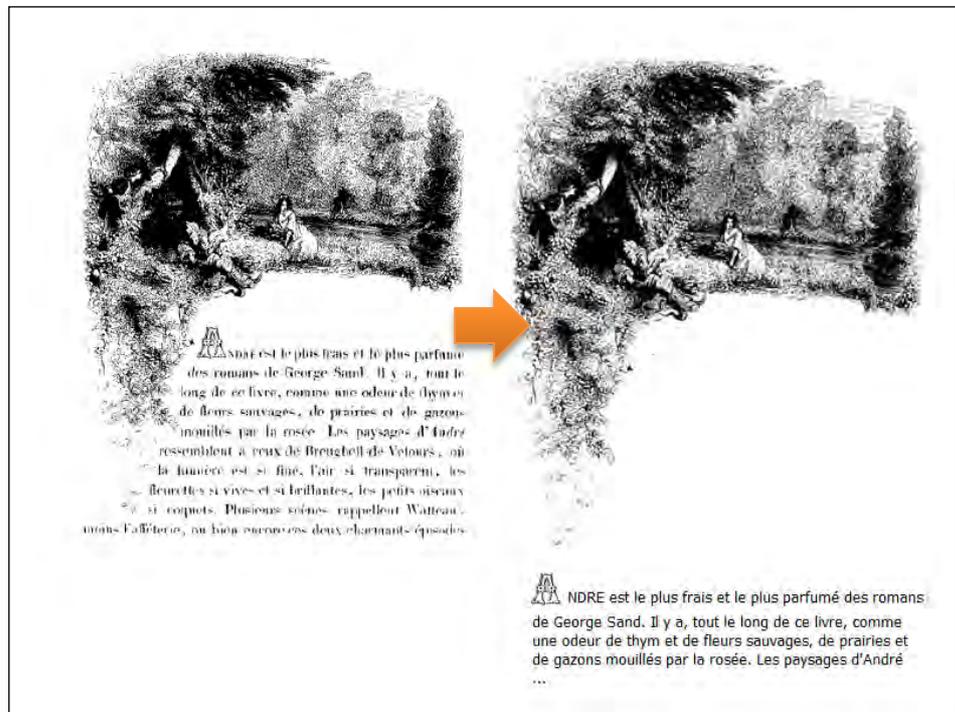
ATTENTION

AUCUN TRAITEMENT DE SUPPRESSION DE TEXTE A L'INTERIEUR DES LETTRES ORNEES NE DOIT ETRE APPLIQUE. PAR CONTRE, LE TEXTE DOIT ETRE TRAITÉ AFIN DE GARANTIR UNE CONTINUITÉ DE LECTURE IMAGE/TEXTE (DANS L'EXEMPLE PRECEDENT, SUR LE MOT « MON »).

Si la lettrine est incluse dans un paragraphe habillée par une illustration, la lettrine et le texte doivent être rejetés après l'illustration.



EXEMPLE



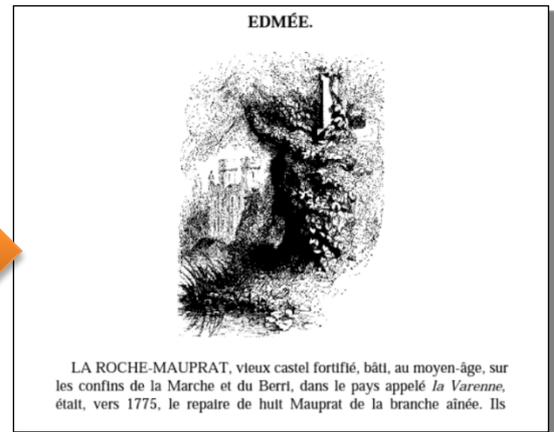
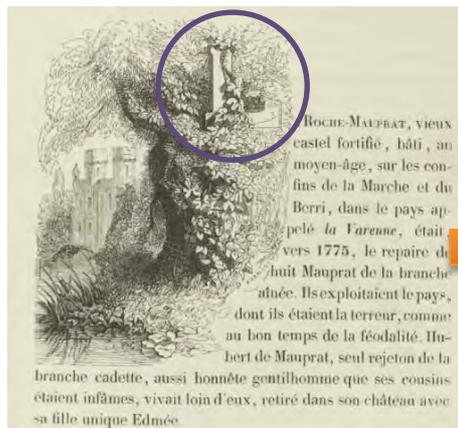
Document

ePub

<http://gallica.bnf.fr/ark:/12148/btv1b8600206q/f243.image>

Si la lettrine est incluse dans une illustration, elle ne peut être extraite afin d'être incluse au début du paragraphe. Dans ce cas, l'illustration est conservée en l'état et la lettrine est gérée en mode texte.

 **EXEMPLE**



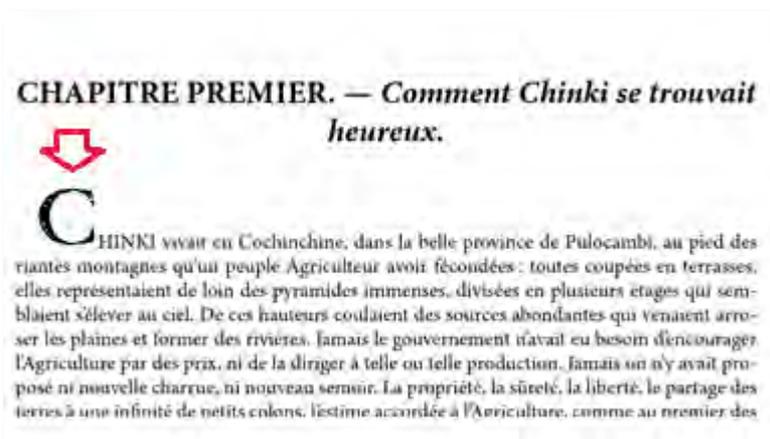
Document

ePub

<http://gallica.bnf.fr/ark:/12148/btv1b8600206q/f33.image>

 **ATTENTION**

LES SIMPLES LETTRINES (SANS ORNEMENT) N'ONT PAS A ETRE TRAITÉES EN MODE IMAGE. ELLES PEUVENT ETRE GÉRÉES EN MODE TEXTE. L'EXEMPLE SUIVANT EST DONC INCORRECT.



<http://gallica.bnf.fr/ark:/12148/bpt6k5540197z/f8.image>

Annotations manuscrites

Les annotations manuscrites ne doivent pas être insérées dans le contenu de l'ePub.



Dans les fichiers ALTO, les annotations sont des GraphicalElement avec un attribut TYPE="manuscrit".

5.3.5 Tableaux

On s'efforcera de composer les tableaux en mode texte. Le prestataire s'accordera avec la BnF sur la mise en forme des tableaux qu'il est en mesure de réaliser. Par exemple :

- les tableaux au-delà de 3 colonnes seront traités en mode image,
- les tableaux jusqu'à 3 colonnes seront traités en mode texte, sauf les tableaux à mise en forme complexe, les tableaux de nombres, les tableaux composés avec un corps de caractères très faible, etc.

Dans cet esprit, les tables et autres listes (de figures, de schémas, de planches, etc.) typées en tableau dans les fichiers ALTO seront produites en mode texte sous la forme d'un tableau HTML à deux colonnes sans bordure, les points de conduite pouvant être supprimés.



Conférences de l'Académie royale de Peinture et de Sculpture.....	1 vol.
La Sculpture en Europe.....	1 vol.
La Sculpture au Salon de 1873.....	1 vol.
La Sculpture au Salon de 1874.....	1 vol.
La Sculpture au Salon de 1875.....	1 vol.
La Sculpture au Salon de 1876.....	1 vol.
La Sculpture au Salon de 1877.....	1 vol.
La Sculpture au Salon de 1878.....	1 vol.
La Sculpture au Salon de 1879.....	1 vol.
La Sculpture au Salon de 1880.....	1 vol.
La Sculpture aux Salons de 1881, 1882, 1883 et à l'Exposition nationale de 1883.....	1 vol.
Hippolyte Flandrin, les Frises de Saint-Vincent-de-Paul. (Conférences populaires).....	1 vol.

Conférences de l'Académie royale de Peinture et de Sculpture	1 vol.
La Sculpture en Europe	1 vol.
La Sculpture au Salon de 1873	1 vol.
...	

5.3.6 Formules

Les formules de mathématiques et de chimie sont à produire en mode image. Elles seront centrées sur la ligne.

Il s'agit de vérifier, maintenant, que les équations (4), accompagnées des relations (8), sont au contraire suffisantes pour établir l'équilibre d'élasticité d'une partie quelconque Ω d'un milieu solide. Multiplions la première équation (4) par $dx dy dz$, et intégrons dans toute l'étendue de Ω , il viendra

$$(10) \left\{ \begin{aligned} & \int \int \int \frac{dN_1}{dx} dx dy dz + \int \int \int \frac{dT_3}{dy} dy dz dx \\ & + \int \int \int \frac{dT_2}{dz} dz dx dy + \int \int \int \rho X_0 dx dy dz = 0. \end{aligned} \right.$$

Dans l'intégrale triple en N_1 , on peut effectuer l'intégration en x , ce qui donnera

$$\int \int dy dz (N_1 - N_1'),$$

Les formules de mathématiques insérées dans le texte d'un paragraphe seront produites à l'aide d'opérateurs mathématiques, de caractères grecs, etc. pour les simples formules, ou insérées en mode image dans le texte du paragraphe pour les formules complexes (fractions, racines, matrices, etc.) :

Cela posé, le tétraèdre devant être en équilibre, sous l'action des forces élastiques qui s'exercent sur ses quatre faces, et des forces qui sollicitent sa masse, les sommes des composantes de ces forces, estimées suivant chaque axe, devront être nulles. A la somme des composantes suivant l'axe des x , la face inclinée fournira le terme $X \varpi$; la face a , le terme $-N_1 \cdot m \varpi$; la face b , $-T_3 \cdot n \varpi$; la face c , $-T_2 \cdot p \varpi$; les forces qui agissent sur la masse donneront un terme égal à ρX_0 multiplié par le volume du tétraèdre, qui est un infiniment petit du troisième ordre; ce cinquième terme disparaîtra donc à la suite des quatre autres, qui sont des infiniment petits du second ordre. Égalant la somme trouvée à zéro, et divisant par ϖ , on obtient la première des équations :



Si un document présente de nombreuses formules insérées dans le texte d'un paragraphe, le prestataire pourra rejeter le document, après accord de la BnF.

5.3.7 Publicités

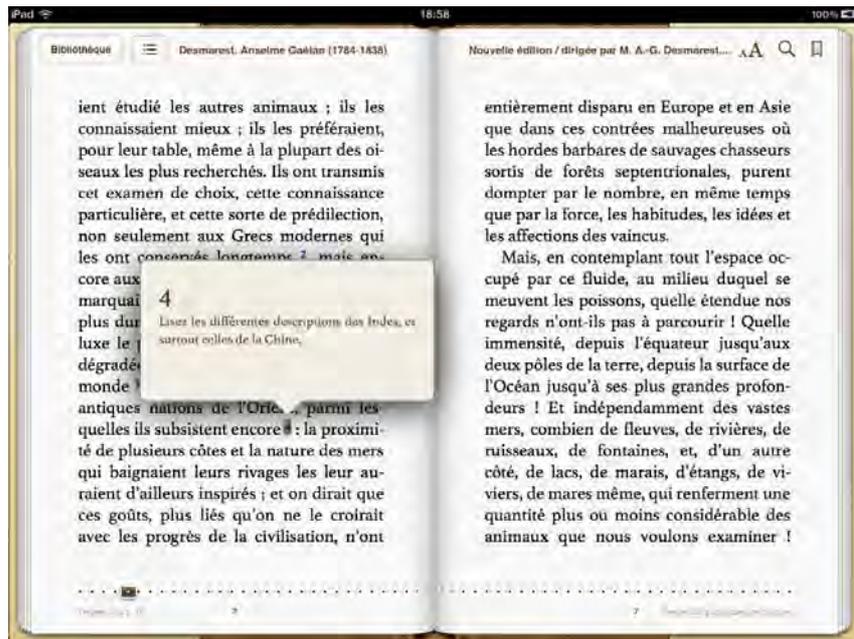
Les publicités et les catalogues d'éditeur sont à reproduire en mode image (cf. aussi section 5.7).

5.3.8 Notes

L'appel de note est repéré dans le texte par une typographie spécifique. Il référence une note dont le texte est généralement placé en bas de page.

Dans l'ePub, le rendu graphique des notes sera assuré par le mécanisme `epub:type="noteref"`.

EXEMPLE



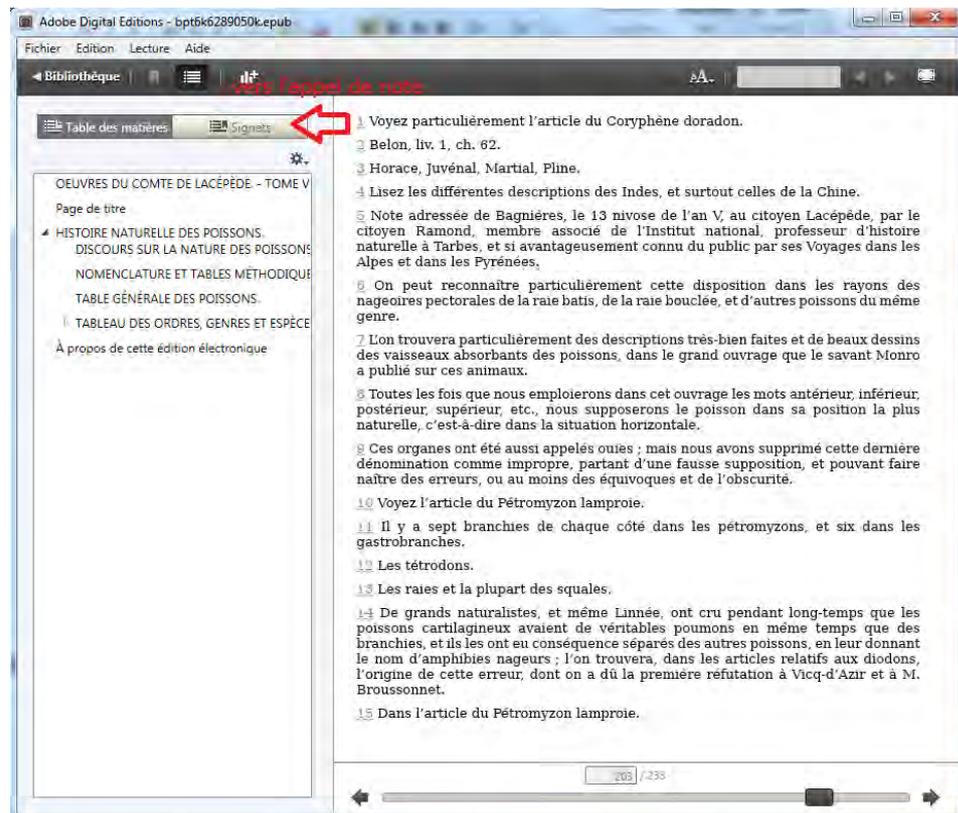
Note epub 3 dans lecteur iBooks

Cependant, afin d'assurer une compatibilité descendante avec la version ePub 2, on veillera à garantir un codage HTML permettant l'affichage des notes par les dispositifs ne sachant pas interpréter le mécanisme epub:type :

- Un lien hypertexte est posé sur l'appel de note.
- Les textes des notes sont positionnés en fin d'ouvrage dans une page HTML dédiée, avec un lien retour posé sur le numéro de note et pointant vers l'appel de note. Cette page est placée avant la page des mentions utiles, cf. section 6.2),.
- Une entrée pointant vers cette page est créée dans la table de navigation de l'ePub, avec pour titre « Notes ».

ATTENTION

CETTE ENTREE N'EST INSEREE DANS LA TABLE DE NAVIGATION QUE SI LE DOCUMENT POSSEDE DES NOTES.



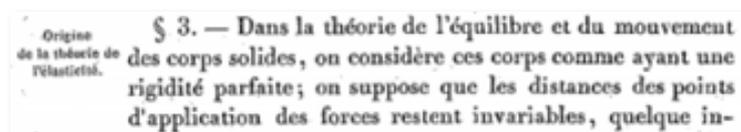
Note epub 2 dans lecteur Adobe Digital Editions

Il convient de noter que :

- Le texte d'une note dans le document original peut couvrir sur plusieurs pages, mais la note dans l'ePub est traitée en un appel et une note.
- Toutes les notes du document original doivent être traitées.
- Les notes sont numérotées de 1 à n à l'échelle du document, aucun label original n'étant conservé (tout est retranscrit en numérotation numérique). Mais si des renvois vers des numéros de note sont présents dans le corps du texte, ce principe ne sera pas suivi et on conservera le mécanisme de numérotation du document.

Cas des notes en marge

Les notes marginales sont placées en regard du texte auquel elle se rapporte. Elles ne sont généralement pas référencées par un appel de note.



Les notes en marge font partie du contenu principal et apparaissent en général dans le flux ALTO dans le PrintSpace, après le corps du texte et les notes de bas de page.

Le rendu graphique des notes en marge sera également assuré par le mécanisme epub:type. Un symbole particulier portant l'appel de note sera placé en tête du paragraphe en regard de la note marginale.



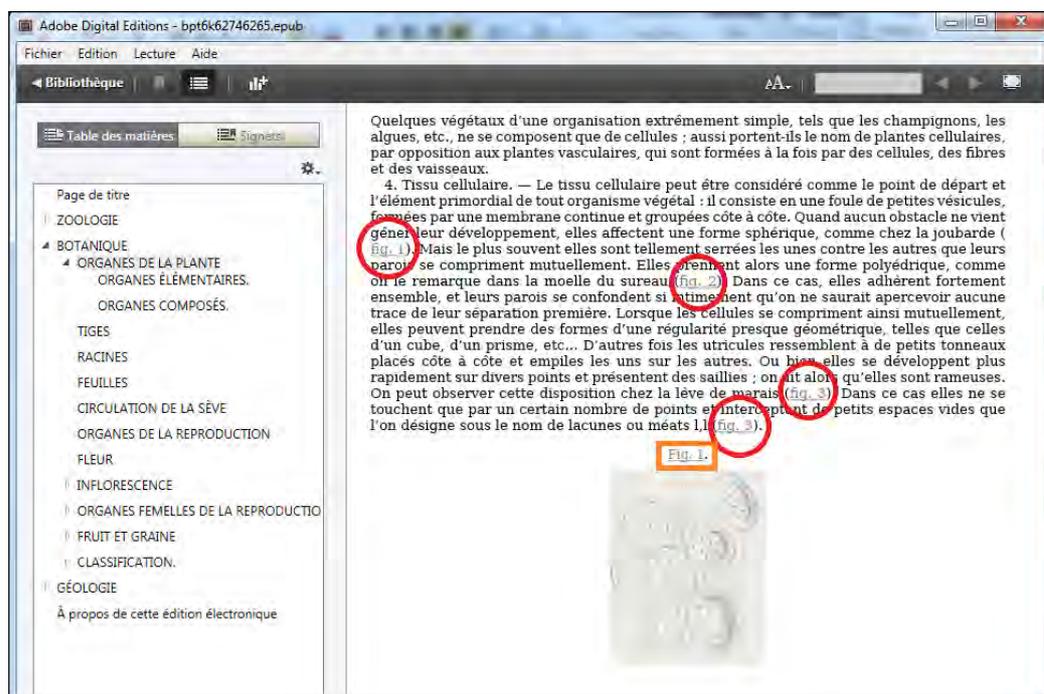
LE PRESTATAIRE VEILLERA A ASSURER UNE COMPATIBILITE TECHNIQUE DESCENDANTE AVEC LA VERSION EPUB 2.

5.3.9 Références croisées

Des liens hypertextes doivent être posés entre certaines composantes du document reliées par un référencement croisé :

- appel de note et note,
- référence de figure et figure,
- référence de tableau et tableau.

L'exemple suivant montre le cas de références de figure.



Figures et références de figure

5.3.10 Divers

Texte en langue étrangère

Les textes dans une langue étrangère écrite avec l'alphabet latin seront reproduits en mode texte ; ils font partie du périmètre du taux qualité (cf. section 5.1).

Les textes écrits avec l'alphabet grec seront reproduits en mode texte ; ils font également partie du périmètre du taux qualité.

Les textes écrits avec un alphabet non latin (cyrillique, langues asiatiques, arabe, hébreu, etc.) seront reproduits en mode image.

Texte manuscrit

Les textes manuscrits présents dans le document original (annotation, lettre manuscrite, etc.) seront reproduits en mode image.

Texte illisible

Les textes (caractères, mots, zones) qui ne seraient pas lisibles par un œil humain sans effort d'interprétation excessive ou qui seraient manquants (page déchirée ou coupée), sont exclus du périmètre du taux qualité et restitués avec une mise en forme spécifique, afin d'informer le lecteur de l'incertitude de leur transcription.

Cette mise en forme sera étudiée de concert par le prestataire et la BnF.

Guillemets répétés

Quand une réplique ou une citation comporte plusieurs alinéas, les alinéas supplémentaires sont parfois introduits par un guillemet ouvrant ou fermant. Il faut supprimer ces guillemets lors de la production de la version numérique du texte.



EXEMPLE

74

GUSTAVE

pouvoir opposer d'illustres ancêtres aux prétentions de la comtesse; pour la première fois il ressentit un véritable chagrin de n'avoir pas obtenu l'avancement qu'il avait sollicité d'après les conseils de M. de C***; il réfléchit que ce dernier avait eu raison plus qu'il ne pensait, et que les sentimens les plus désintéressés pouvaient faire un besoin des hochets de la vanité pour l'homme enclavé dans les institutions tyranniques de la société. « Si j'avais réussi, se » disait-il, j'aurais été bien loin encore » d'être le gendre que madame de P*** » eût choisi, mais du moins la carrière » m'eût été ouverte, et avec quel zèle » infatigable ne m'y serais-je pas élancé, » si j'avais entrevu dans l'avenir la main » de Laure pour prix de mes travaux? » Mais jamais cette inestimable con-

<http://gallica.bnf.fr/ark:/12148/bpt6k5542534f/f89.image>

5.4 Couverture et page de titre

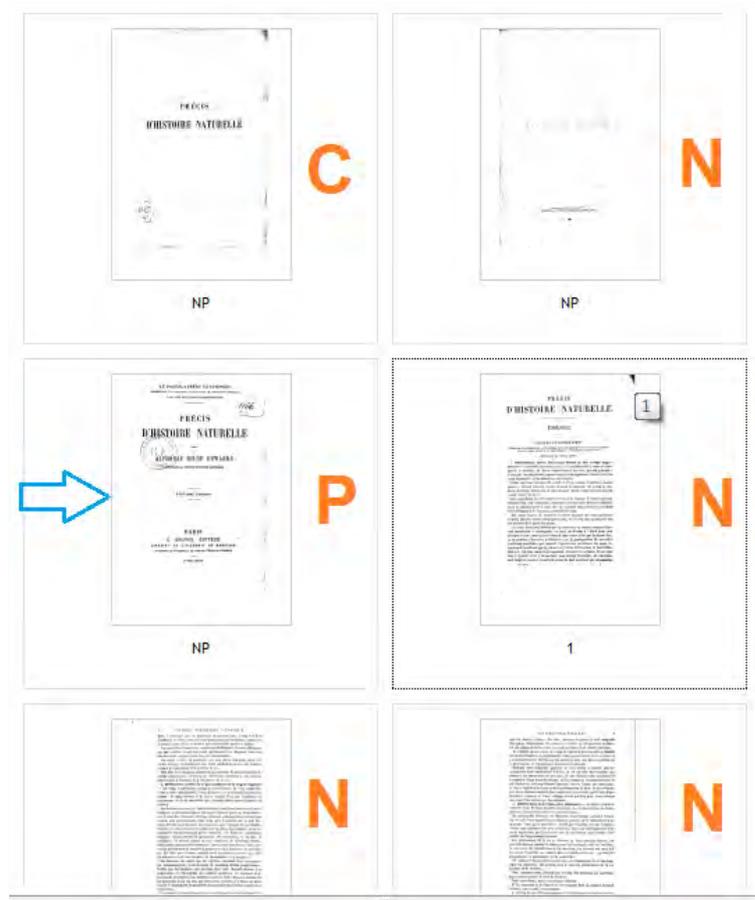
5.4.1 Couverture

Une couverture doit obligatoirement être créée.

Choix de la couverture

Il n'est pas prévu de conserver la couverture du document, qui est souvent une couverture ajoutée ultérieurement à la publication de l'œuvre. La couverture de l'ePub sera donc la page de titre du document original, telle que typée dans le manifeste numérique du document. La figure suivante illustre ce principe.

EXEMPLE



<http://gallica.bnf.fr/ark:/12148/bpt6k370198q/f1.planchecontact.r=.langFR>

La page de titre apparaîtra deux fois dans l'ePub : une fois en mode image (la couverture), une fois en mode texte (dans le corps de l'ouvrage).

NOTES

Le manifeste refNum prévoit les typages suivants :

- page de couverture : typePage=C
- page de titre : typePage=P
- page d'entrée : typePage=E
- page normale : typePage=N
- page illustrée : typePage=D
- page de table des matières : typePage=T
- page de table d'index : typePage=I

En l'absence de page typée P dans le manifeste refNum, la page portant le plus d'informations bibliographiques doit être sélectionnée en tant que page de titre (et donc page de couverture). C'est généralement la page typée E.

Dans le cas d'un journal, la page de titre et la page de couverture sont confondues et elle est identifiée P dans le refNum.



<http://gallica.bnf.fr/ark:/12148/bpt6k6251642s/f1.planchecontact.r=le%20rire.langFR>

Dans le cas de certains *périodiques*, la page de titre sera la page de table des matières du document original. Si la page de titre est aussi une page de table des matières ou d'index, c'est le typage en table ou en index qui prime dans le refNum, soit T ou I.



<http://gallica.bnf.fr/ark:/12148/bpt6k6200173x/f1.planchecontact>

Dans ce cas également, la page de titre apparaîtra deux fois dans l'ePub : une fois en mode image (la couverture), une fois en mode texte (dans le corps de l'ouvrage).

Certaines œuvres présentent plusieurs pages de titre, le titre de l'œuvre et le titre du ou des tomes inclus dans l'œuvre. Dans ce cas, il faut prendre comme page de couverture de l'ePub la page de titre du volume (dans l'exemple suivant, la page 1).



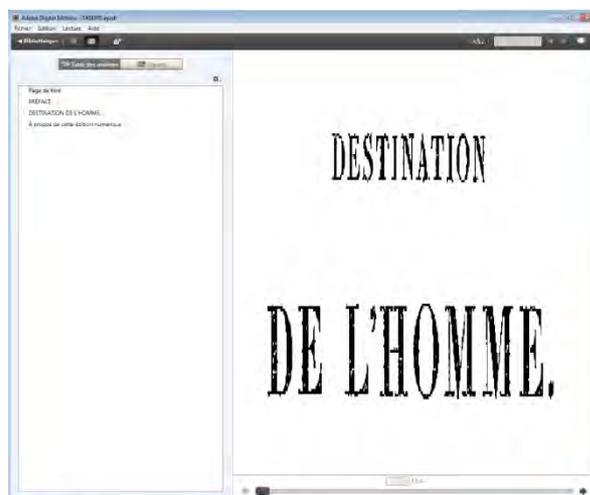
<http://gallica.bnf.fr/ark:/12148/btv1b20000797/f1.planchecontact>

Format de la couverture

L'image qui forme la couverture de l'ePub sera recadrée de façon à former un rectangle correct et fidèle au format de l'ouvrage d'origine.



L'IMAGE NE SERA PAS RECADREE AUTOUR DU TEXTE, FAUTE DE QUOI, LE RESPECT DU FORMAT D'ORIGINE NE SERAIT PLUS ASSURE.



Couverture artificiellement recadrée sur le titre

L'image sera redressée si nécessaire afin d'assurer que les lignes de texte présentes sur la couverture soient horizontales (une tolérance maximale de 0,25° sera admise).

Sa taille sera suffisante pour un affichage de qualité sur ordinateur ou tablette moderne (sa largeur minimale sera de 800 pixels).

Quel que soit le dispositif de lecture utilisé (liseuse, tablette, logiciel de lecture), la couverture doit :

- être centrée dans la fenêtre d'affichage ;
- occuper la totalité de la fenêtre d'affichage (hors les marges éventuellement ajoutées par le dispositif de lecture) ;
- se redimensionner en fonction de l'orientation du dispositif de lecture ou de la taille de la fenêtre d'affichage (dans le cas d'un logiciel de lecture) ;
- apparaître dans la bibliothèque du dispositif, sous la forme d'une vignette (et non sous la forme de la couverture par défaut du dispositif de lecture).

Le mode colorimétrique de l'image de couverture restituée dans l'ePub est défini par les règles suivantes :

Image dans le document original	Images de la couverture dans l'ePub	Commentaires
Couleurs	couleurs	
niveaux de gris	niveaux de gris	
noir et blanc	niveaux de gris	Les niveaux de gris sont obtenus lors du redimensionnement de l'image.

5.4.2 Page de titre

La page de titre constitue le premier contenu de l'ePub. Elle doit être mise en page dans l'ePub conformément à la charte graphique.

La page de titre est soit la page typée P dans le manifeste refNum, soit la page portant le plus d'informations bibliographiques (cf. section précédente). Elle doit être insérée dans l'ePub en mode texte. Si elle contient des éléments graphiques, ils seront gérés en mode image.

Comme il a été dit précédemment, la page de titre apparaît deux fois dans l'ePub : une fois en mode image (la couverture), une fois en mode texte (dans le corps de l'ouvrage).



Couverture



Page de titre

Place de la page de titre dans le livre numérique

La page de titre est insérée en tout début d'ePub, et une entrée est créée au début de la table de navigation de l'ePub, avec le texte suivant :

Page de titre



Les pages de plats, de garde et de faux titre doivent être supprimées.

5.5 Tables de navigation

5.5.1 Table des matières

Une table des matières conforme à la norme ePub 3 est créée à partir du contenu du document :

```
<html xmlns="http://www.w3.org/1999/xhtml"
xmlns:epub="http://www.idpf.org/2007/ops">
<head> <title>Oeuvres...</title></head>
```

```
<body>
<h1>Table des matières</h1>
```

```

<ol>
  <li id="toc1"><a href="../Text/ebpt6k6289050k_fm01.html">OEUVRES DU COMTE DE
  LACÉPÈDE. - TOME V.</a></li>

  <li id="toc2"><a href="../Text/ebpt6k6289050k_tp01.html">Page de titre</a></li>

  <li id="toc3">
    <a href="../Text/ebpt6k6289050k_p01.html">HISTOIRE NATURELLE DES
    POISSONS.</a>
    <ol>
      <li id="toc4"><a href="../Text/ebpt6k6289050k_c01.xhtml">DISCOURS SUR LA
      NATURE DES POISSONS.</a></li>
      <li id="toc5"><a href="../Text/ebpt6k6289050k_c02.xhtml">NOMENCLATURE ET
      TABLES MÉTHODIQUES DES POISSONS.</a></li>
    </ol>
  </li>
  ...

```

La structure de cette table est basée sur la table des matières ou le sommaire du document originel. Si ce dernier ne dispose pas de table des matières ou de sommaire, une structure sera déduite de la structuration logique du contenu. Si le contenu n'est pas structuré (par exemple un roman sans découpage en chapitres), aucune entrée de navigation ne sera insérée dans la table.



Cette table est créée même si le document original ne comporte pas de sommaire ou de table des matières, car elle doit comporter obligatoirement des liens vers :

- la page de titre (cf. section 5.4.2).
- la page comportant les mentions utiles (cf. section 6.2.1).

Dans le cas où cette table doit être créée ex-nihilo, on veillera à ne pas y insérer les éléments qui sont traditionnellement absents des sommaires et tables des matières : dédicace, remerciement, pages liminaires (copyright), publicité et catalogue d'éditeur, achevé d'imprimé, etc.

La table doit respecter les critères suivants :

- Comporter au maximum quatre niveaux de titres. Si ce critère conduit à une table très volumineuse, le prestataire pourra réduire le nombre de niveaux de la table. Il pourra notamment ne pas inclure les sous-niveaux numérotés (I, II, III..., 1, 2, 3...) :

Titre de chapitre

I

II

III

...

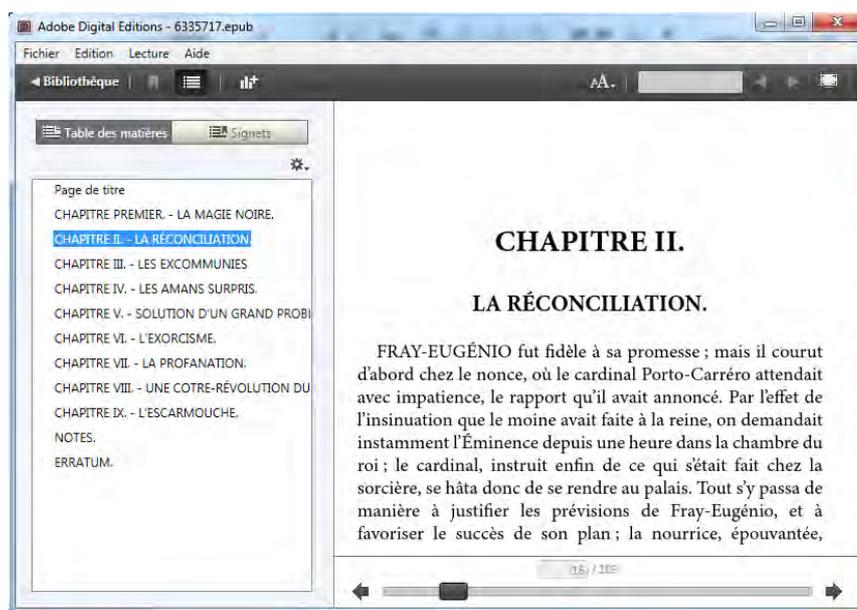
Inclure cependant ces niveaux dès lors qu'ils incluent un élément textuel :

Titre de chapitre

I – **Sous-titre**

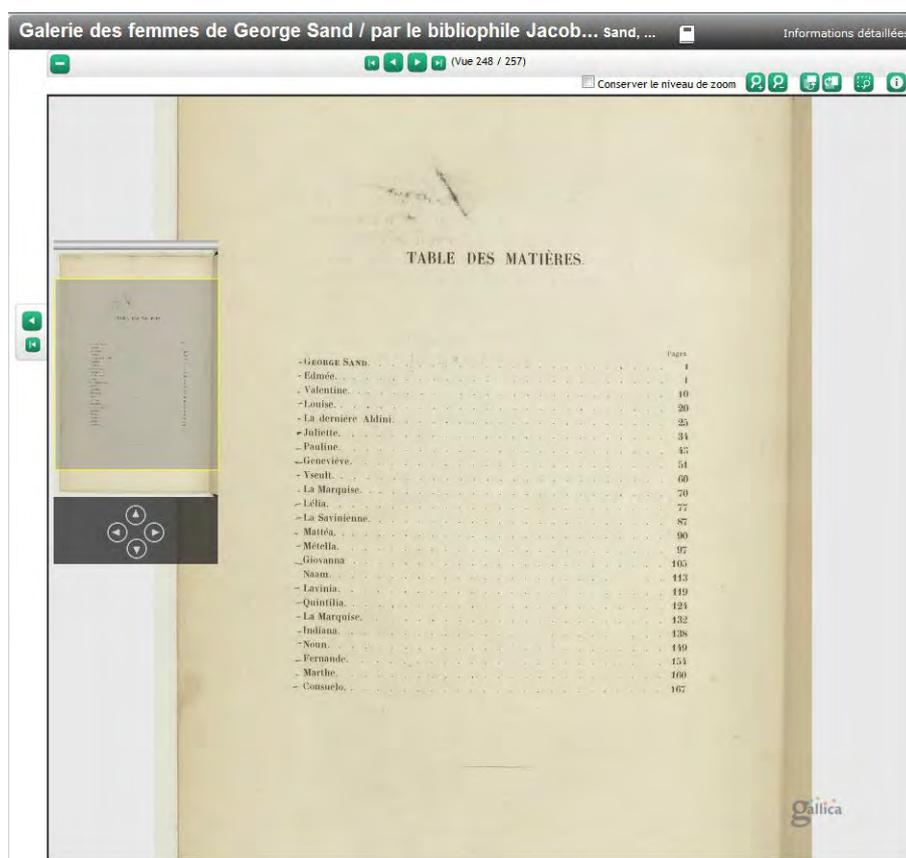
...

- Regrouper numérotation et titre (en les séparant par un tiret), lorsqu'ils sont composés sur deux lignes différentes dans le document d'origine :



Le sommaire ou la table de matières de l'original sont supprimés du contenu du livre numérique (pages typées T ou I dans le refNum, sauf exception : quand la page T ou I est la page de titre de facto).

L'exemple suivant montre une TDM qui ne doit pas être conservée dans l'ePub.



<http://gallica.bnf.fr/ark:/12148/btv1b8600206q/f248.image>

On ne conserve dans le contenu du livre numérique que les tables des matières ou sommaires qui ne comportent pas de renvois de page dans l'original. Ces tables des matières (ou sommaires) doivent être reproduites en tant que texte brut, *sans liens hypertextes* (pages typées R dans le refNum). Elles doivent être insérées en tête de l'ePub (après la couverture et la page de titre), même si elles sont placées en fin de l'ouvrage d'origine.

Les pages d'index et de table des matières déqualifiées (cf. Référentiel de traitement des tables, pages typées Y et Z dans le refNum) doivent également être reproduites en tant que texte brut, sans liens hypertextes. Elles doivent être insérées à un emplacement équivalent à celui de l'ouvrage d'origine.



Afin d'assurer une compatibilité descendante avec la version ePub 2, on veillera à produire la table de navigation au format .ncx, en plus de la table de navigation au format HTML de la version ePub 3.

5.5.2 Table des repères

Une table dite « des repères » (*landmarks*) doit être présente dans l'ePub. Elle permet d'identifier

- les principaux éléments présents dans un ouvrage (préface, début du contenu principal, annexes, etc.)
- les outils de navigation de l'ePub (table des matières, table des figures, index, etc.).

```

<nav epub:type="landmarks">
  <h1>Repères</h1>
  <ol>
    <li>
      <a epub:type="toc" href="tdm.xhtml">
        Table des matières</a>
    </li>
    <li>
      <a epub:type="bodymatter" href="chap001.xhtml#bodymatter">
        Chapitre 1</a>
    </li>
    ...
    <li>
      <a epub:type="index" href="index.xhtml">
        Index</a>
    </li>
  </ol>
</nav>

```



La table des repère doit être structurée sur un seul niveau hiérarchique.

Les éléments à insérer dans la table des *landmarks* sont constitués d'un sous-ensemble des éléments identifiés dans les phases de production précédente. Le tableau suivant en présente une liste non exhaustive.

Entrées de la table	A insérer
Début du contenu principal	obligatoire
Table des matières	obligatoire
Liste des illustrations	optionnel
Liste des tableaux	optionnel
Index	optionnel
Glossaire	optionnel
Bibliographie	optionnel

5.5.3 Table des numéros de page

Le marquage des numéros de page de l'œuvre physique consiste à insérer une référence aux folios du document d'origine au sein des contenus numériques.



Le marquage des numéros de page de l'œuvre physique est une bonne pratique en matière d'accessibilité, puisqu'il permet à des lecteurs sur supports papier et numérique de communiquer entre eux afin de faire référence à une localisation donnée de l'œuvre. Dans le cadre d'une prestation de numérisation, réaliser ce marquage est aisé puisque la numérisation est réalisée à l'échelle de la page : les folios physiques sont donc connus (via les informations du manifeste numérique du document).

Les pages à marquer sont identifiées par le manifeste du document numérique via leur type de pagination (attribut `typePagination` du `refNum`) et leur valeur de numérotation (attribut `numeroPage` du `refNum`) (cf. Référentiel d'enrichissement des métadonnées, sections 3.2 et 3.3).

Le mode de numérotation est donné par les règles suivantes, à appliquer selon le type de pagination :

- N (pages non paginées) : pas d'entrée à insérer dans la table
- A (chiffres arabes) : une entrée en chiffres arabes à insérer dans la table, donnée par la valeur de l'attribut `numeroPage`
- R (chiffres romains) : une entrée en chiffres romains à insérer dans la table, donnée par la conversion en numérotation romaine de la valeur de l'attribut `numeroPage`
- F (cas des paginations doublées, type feuillet) : une entrée à insérer dans la table, donnée par la valeur de l'attribut `numeroPage`
- X (autres cas) : une entrée à insérer dans la table, donnée par la valeur de l'attribut `numeroPage`.

Le marquage consiste à insérer dans le contenu numérique une balise mentionnant le folio de chaque page numérotée du document d'origine. La balise doit être insérée avant le contenu présent en haut de page (du document d'origine).

```
<p>Texte texte texte<span xml:id="page36 1" epub:type="pagenumber">36 1</span>
texte texte texte.</p>
```

Une table dite *page-list* permet de communiquer cette liste de numéros de page au lecteur :

```
<nav epub:type="page-list">
<h1>Pages</h1>
<ol>
<li><a href="pagedetitre.xhtml#pageii">II</a></li>
...
<li><a href="preface.xhtml#pagev">V</a></li>
<li><a href="preface.xhtml#pagevi">VI</a></li>
<li><a href="chapitre001.xhtml#page001">1</a></li>
<li><a href="chapitre001.xhtml#page002">2</a></li>
...
<li><a href="chapitre001.xhtml#page256">256</a></li>
</ol>
```

5.6 Index

Les pages d'index du document d'origine sont à insérer dans l'ePub.

Le taux qualité éditorial ne s'applique pas sur les numéros de page des entrées d'index, lesquels seront traités en OCR brut. Aucun renvoi hypertexte ne sera créé sur ces numéros de page.

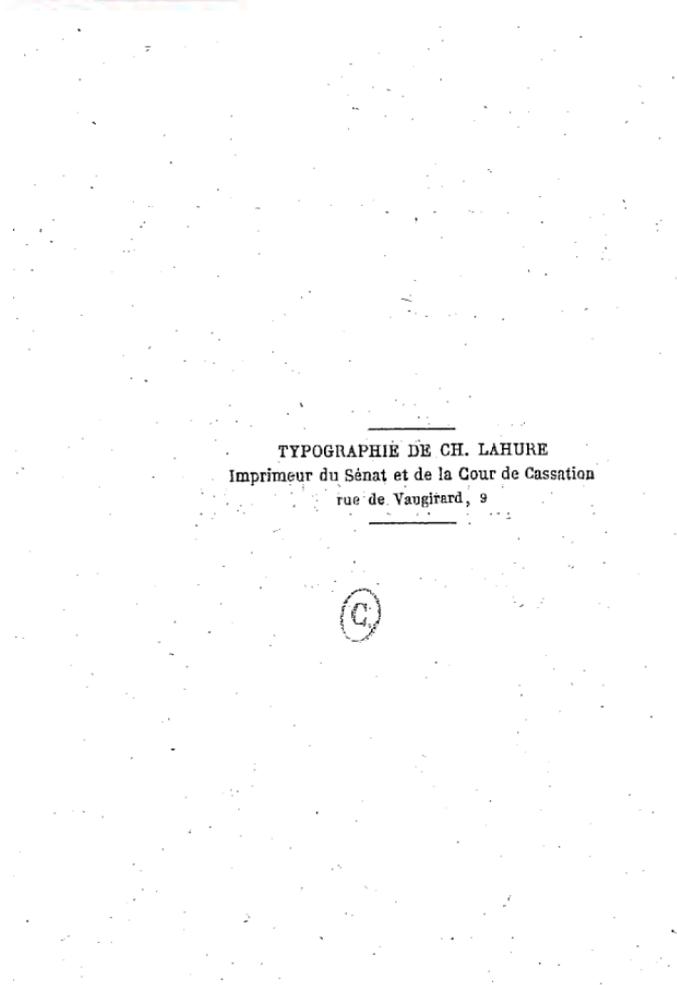
5.7 Pages spéciales de l'original

5.7.1 Reliures

Les images de reliure de l'original doivent être supprimées.

5.7.2 Achevés d'imprimé, mentions d'imprimeur, etc.

Les achevés d'imprimé et autres mentions d'imprimeur doivent être produits en mode texte.



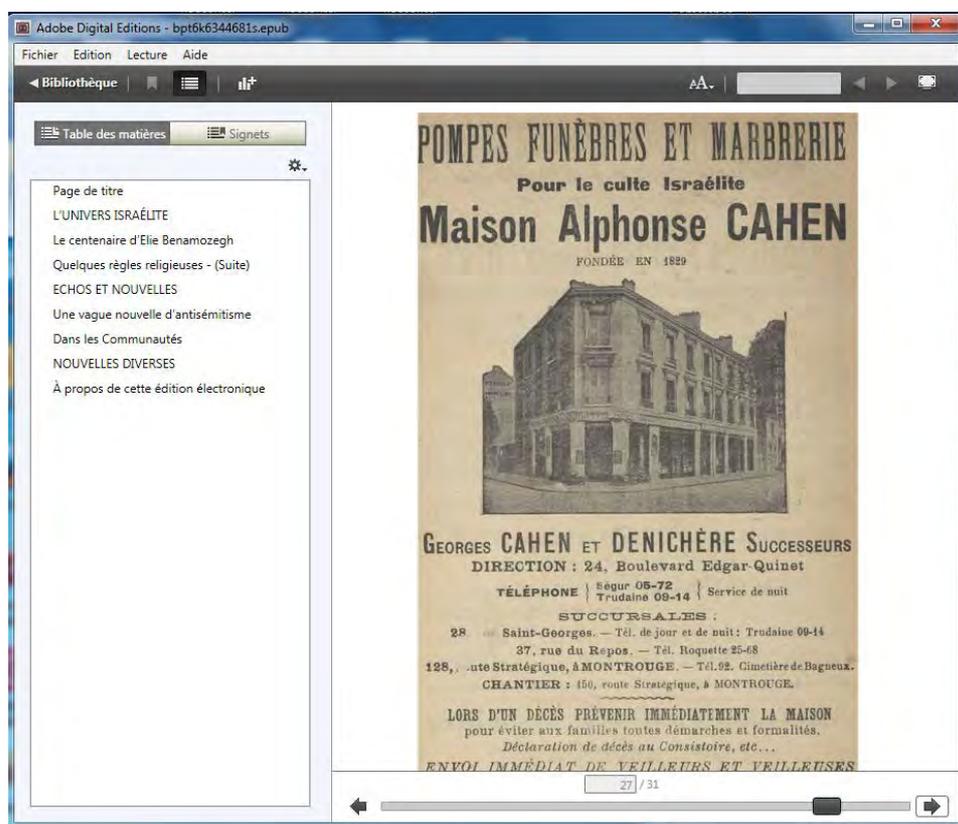
TYPOGRAPHIE DE CH. LAHURE
Imprimeur du Sénat et de la Cour de Cassation
rue de Vaugirard, 9



Exemple de mention d'imprimeur

5.7.3 Publicités

Les autres pages spéciales comme (liste non exhaustive) publicité (cf. exemple ci-dessous), catalogue d'éditeur (pages typées A dans le refNum), seront produites en mode image, sauf accord particulier avec le prestataire.



Exemple de page de publicité

5.7.4 Errata

Les pages d'errata seront produites en mode texte.

5.7.5 Tables

Les tables des illustrations, des schémas, des planches, etc. seront produites en mode texte (cf. section 5.3.5). Les renvois hypertextes vers les contenus pointés par les numéros de page des tables seront créés.



Les renvois hypertextes pourront ne pas être créés pour les tables volumineuses (supérieures à 100 entrées).

6. METADONNEES ET MENTIONS UTILES

6.1 Métadonnées bibliographiques

Les métadonnées bibliographiques sont obtenues par le prestataire selon deux voies :

- Dans le cas d'un marché de réfection, les métadonnées sont récupérées du refNum, puisque celui-ci a été créé et enrichi en amont de la fabrication des ePub, lors d'une prestation antérieure. Les métadonnées peuvent également être obtenues dans l'OAI NUM de la BnF.
- Dans le cas d'un marché intégré (numérisation puis production ePub), les métadonnées sont fournies dans le bordereau de traitement aller (BTA) et doivent être copiées comme telles. Les métadonnées peuvent également être obtenues dans l'OAI CAT de la BnF.

6.1.1 Identifiant

On ajoute un élément dc:identifiant permettant d'identifier de manière unique le livre numérique :

```
<dc:identifiant id="dcidid" opf:scheme="URI">  
    ark:/12148/bpt6k54272182.epub  
</dc:identifiant>
```

L'ark de l'exemplaire numérique est calculé par le prestataire à partir de la formule fournie par la BnF, conformément au document BnF-ADM-2007-013608-02 : « Comment créer des identifiants ark pour des tests (SPAR SAIA) »

Dans le cas d'une œuvre issue d'une œuvre complète :

```
<dc:identifiant id="dcidid" opf:scheme="URI">  
    ark:/12148/bpt6k54272182/w1.epub  
</dc:identifiant>
```

où « w1 » signifie que l'on parle de la première œuvre du document (le numéro à utiliser est celui du bordereau de remise).

Le type d'identifiant sera décrit, ainsi que son organisme gestionnaire :

```
<meta refines="#pub-id" property="identifiant-type" scheme="xsd:URI">ark</meta>  
<meta refines="#pub-id" property="meta-auth">www.cdlib.org</meta>
```



Pour identifier une version spécifique d'un ePub (par exemple dans le cas d'un marché de réfection d'ePub produits antérieurement), un identificateur sera construit en combinant l'identificateur unique avec la dernière date de modification de la publication :

```
<dc:identifiant id="pub-id">ark:/12148/bpt6k6289050k.epub</dc:identifiant>  
<meta property="dcterms:modified">2013-05-01T12:00:00Z</meta>
```

6.1.2 Titre

La métadonnée dc:title contient le titre de l'œuvre (cf. balise <titre> du manifeste refNum).



La langue d'un titre peut être caractérisée avec l'attribut xml:lang :

```
<dc:title id="t1" xml:lang="la">M. T. Ciceronis Brutus, sive de Claris oratoribus</dc:title>
```

ePub 3 enrichit la description des publications avec la propriété title-type. Cette propriété peut prendre les valeurs suivantes :

- main : titre principal
- subtitle : sous-titre
- short : titre court, titre usuel
- collection : collection, série
- edition : édition
- expanded : titre complet

Monographie

Les valeurs main, subtitle, edition de la propriété title-type sont adaptées à la description des monographies. Elles sont associées aux métadonnées obtenues dans l'OAI, seul le titre étant mentionné dans le refNum :

- sous-titre : champ 245\$e
- édition : champ 250\$a



<http://gallica.bnf.fr/ark:/12148/bpt6k6289050k>

Catalogue :

```
000 n0 m 22 a
001 FRBNF307069690000002
008 970701m 1826 1833 frfre b
009 a
100 $312027512 $w 0 b....$aLacépède$mÉtienne de$d1756-1825$40070
245 1 $aOeuvres du comte de Lacépède,...$dTexte imprimé
250 $aNouvelle édition$dirigée par M. A.-G. Desmarest,...
```

ePub :

```
<dc:title id="t1">Œuvres du comte de Lacépède,... </dc:title>
<meta refines="#t1" property="title-type">main</meta>
```

```
<dc:title id="t2">Nouvelle édition / dirigée par M. A.-G. Desmarest,...</dc:title>
```

```
<meta refines="#t2" property="title-type">edition</meta>
```

On souhaite imposer que les dispositifs de lecture utilisent le titre complet plutôt que le titre principal (main), car le titre complet est plus riche (notamment pour les multivolumes et les périodiques). Pour ce faire, il faut utiliser :

- la valeur `extended` de la propriété `title-type` qui permet de décrire le titre complet ;
- la propriété `display-seq` en spécifiant l'identifiant du titre complet comme étant à afficher en priorité.

L'exemple précédent devient donc :



```
<dc:title id="t1">Œuvres du comte de Lacépède,... </dc:title>  
<meta refines="#t1" property="title-type">main</meta>
```

```
<dc:title id="t2">Nouvelle édition / dirigée par M. A.-G. Desmarest,...</dc:title>  
<meta refines="#t2" property="title-type">edition</meta>
```

```
<meta id="t3"> Œuvres du comte de Lacépède,... – Nouvelle édition / dirigée par  
M. A.-G. Desmarest,...</dc:title>  
<meta refines="#t3" property="title-type">extended</meta>  
<meta refines="#t3" property="display-seq">1</meta>
```

Œuvre multivolume

La propriété `group-position` permet de décrire le rang de la publication au sein de l'œuvre.

La valeur `extended` de la propriété `title-type` permet de décrire le titre complet avec sa tomaison/volumaison.

Les informations de tomaison/volumaison sont stockées dans la balise `<tomaison>` du `refNum` (cf. Référentiel d'enrichissement des métadonnées) ou dans l'OAI. On peut trouver trois mentions hiérarchiques au maximum.

Ces informations de tomaison/volumaison seront insérées après le titre, avec un séparateur tiret, selon la syntaxe suivante :

titre - tomaison1[, tomaison2][, tomaison3]



Une mention de tomaison/volumaison est toujours un couple type/valeur. Dans le `refNum`, la balise `type` est normée et ne peut contenir que les mentions suivantes, qu'il faudra convertir selon ces règles :

Valeur refNum	Transcription ePub	Exemples
Année	<i>tel que</i>	Nouveau Mercure galant - 1714
Edition	ed.	

Fascicule	fasc.	
Numéro	n°	La Chronique des arts et de la curiosité - n° 8
Partie	partie	
Section	section	
Série	<i>tel que</i>	Scènes de la vie privée - Béatrix
Tome	t.	Les Misérables - t. 2
Volume	vol.	Mademoiselle de La Seiglière - vol. 1



<http://gallica.bnf.fr/ark:/12148/bpt6k55450567>

RefNum :

```
<tomaison>
  <type>Tome</type>
  <valeur>2</valeur>
</tomaison>
```

ePub :

```
<dc:title id="t1">Les Misérables</dc:title>
<meta refines="#t1" property="title-type">main</meta>
<meta refines="#t1" property="group-position">2</meta>
```

```
<dc:title id="t2">Les Misérables - Tome 2</dc:title>
<meta refines="#t2" property="title-type">extended</meta>
<meta refines="#t2" property="display-seq">1</meta>
```



On peut indiquer une variante favorisant le classement alphabétique des titres avec la propriété `file-as` :

```
<dc:title id="t2">Les Misérables - Tome 2</dc:title>
<meta refines="#t2" property="title-type">extended</meta>
<meta refines="#t2" property="file-as">Misérables, Les - Tome 2</meta>
```

Périodique

La valeur `extended` de la propriété `title-type` permet de décrire le titre du périodique avec ses informations de `tomaison`.

Dans ce cas également, on priorise l'affichage du titre complet avec la propriété `display-seq`.



<http://gallica.bnf.fr/ark:/12148/bpt6k63197487.image>

RefNum :

```
<tomaison>
  <type>Année</type>
  <valeur>1714</valeur>
</tomaison>
```

ePub :

```
<dc:title id="t1">Nouveau Mercure galant</dc:title>
<meta refines="#t1" property="title-type">main</meta>
```

```
<dc:title id="t2">Nouveau Mercure galant - 1714</dc:title>
<meta refines="#t2" property="title-type">extended</meta>
<meta refines="#t2" property="display-seq">1</meta>
```



<http://gallica.bnf.fr/ark:/12148/bpt6k6200307m.image>

RefNum :

```
<tomaison>
  <type>Numéro</type>
  <valeur>8</valeur>
</tomaison>
```

ePub :

```
<dc:title id="t1">La Chronique des arts et de la curiosité</dc:title>
<meta refines="#t1" property="title-type">main</meta>
```

```
<dc:title id="t2">La Chronique des arts et de la curiosité - n° 8</dc:title>
<meta refines="#t2" property="title-type">extended</meta>
<meta refines="#t2" property="display-seq">1</meta>
```

6.1.3 Auteur

La métadonnée dc:creator spécifie l'auteur principal de l'œuvre (cf. balise <auteur> du refNum).



Quand l'auteur n'est pas connu, utiliser la valeur :

- "Inconnu" pour une monographie (cf. balise <genre>MONOGRAPHIE</genre> du refNum)
- "Collectif" pour un périodique (<genre>PERIODIQUE</genre>)

Plusieurs auteurs peuvent être associés à la publication, avec autant de balises `dc:creator` que d'auteurs. Leurs rôles respectifs sont décrits avec une balise `<meta refines>`. Ces informations sont à chercher dans l'OAI :

- auteurs : champs `100amde` (auteur principal) et `700amde` (autres auteurs),
- rôles : dans la sous-zone \$4 des champs 100 et 700, la valeur 0070 identifie les auteurs.



On indiquera une variante favorisant le classement alphabétique par auteur avec la propriété `file-as` :

```
<dc:creator id="auteur">Étienne de Lacépède (1756-1825)</dc:creator>
<meta refines="#auteur" property="role" scheme="marc:relators">aut</meta>
<meta refines="#auteur" property="file-as">Lacépède, Étienne de</meta>
```

6.1.4 Contributeurs

Des contributeurs peuvent être associés à la publication, avec les balises `dc:contributor` et `meta refines`. Ces informations sont à chercher dans l'OAI :

- contributeurs : champ `700amde`
- rôles : dans la sous-zone \$4 du champ 700, les valeurs autres que 0070 identifient des contributeurs.



<http://gallica.bnf.fr/ark:/12148/bpt6k6289050k>

Catalogue :

700 \$310615544 \$w.0..b.....\$aDesmarest\$mAnselme Gaétan\$d1784-1838\$40360

ePub :

```
<dc:creator id="auteur">Lacépède, Étienne de (1756-1825)</dc:creator>
<meta refines="#auteur" property="role" scheme="marc:relators">aut</meta>
```

```
<dc:contributor id="editeur">Desmarest, Anselme Gaétan (1784-
1838)</dc:contributor>
<meta refines="#editeur" property="role" scheme="marc:relators">edt</meta>
```

6.1.5 Editeur

La métadonnée `dc:publisher` décrit l'éditeur de la publication numérique (l'ePub). Si l'on respecte l'esprit du *Dublin Core Metadata Element Set*, ce devrait être la BnF, mais il a été choisi de décrire l'éditeur de l'œuvre originelle (cf. balise `<editeur>` du `refNum`).



<http://gallica.bnf.fr/ark:/12148/bpt6k6289050k>

<dc:publisher>Ladrange et Verdière (Paris) </dc:publisher>

6.1.6 Dates

Date de publication

La balise dc:date contient la date du jour (soit celle de la création du livre numérique), au format ISO :

AAAA-MM-JJ



CETTE METADONNÉE NE CONTIENT PAS LA DATE D'ÉDITION DU DOCUMENT D'ORIGINE ! (CF. SECTION SUIVANTE).

Date d'édition de l'œuvre originelle

Il n'existe pas de métadonnée de premier rang pour décrire la date d'édition de l'œuvre originelle. On utilisera donc cette formulation pour mentionner la date de publication du document d'origine (dans le refNum : <dateEdition>) :

<meta property="dcterms:issued">date</meta>

La valeur de date sera insérée au format ISO :

AAAA[-MM[-JJ]]



- Quand <dateEdition> contient une plage de dates, prendre la première.
- Quand <dateEdition> contient une mention alphanumérique, extraire la date : « an XI (1802) » devient « 1802 ».

Lorsqu'il n'y a pas de <dateEdition> dans les métadonnées (« s.d. » au catalogue, et pas de date trouvée lors de l'exemplarisation et portée au refNum), la métadonnée <meta property="dcterms:issued">...</meta> n'est pas insérée.



L'ÉLÉMENT VIDE <META PROPERTY="DCTERMS:ISSUED"/> N'EST PAS ACCEPTÉ.



Le bon affichage des métadonnées des ePub 3 dans les bibliothèques des logiciels de lecture de livres numériques (notamment les informations de tomaison et de date) n'est pas assuré à l'heure actuelle (mi 2013). Le prestataire fera donc le point avec la BnF en début de marché, afin d'envisager des solutions de repli. On pourra notamment envisager d'ajouter la date de publication de l'œuvre originelle à la fin de `dc:title`, entre parenthèses :

titre (date d'édition)

6.1.7 Langue

`dc:language` indique la langue principale de l'œuvre. Cette valeur doit être extraite de l'OAI :

- langues : champ 008, positions 31-33 : langues en code ISO
- si positions 31-33= « mis » ou « mul » ou « mmm » : zone 041



Catalogue :

008 970701m 1826 1833 frfre b

ePub :

```
<dc:language>fr</dc:language>
```

Dans le cas d'œuvres multilingues ou traduites, il peut y avoir plusieurs balises :



Catalogue :

008 970701m 1826 1833 frmul b
041 0 \$afre\$alat

ePub :

```
<dc:language>fr</dc:language>
```

```
<dc:language>la</dc:language>
```

6.1.8 Source

On ajoute un élément `dc:source` pour indiquer la ressource numérique dont est dérivé le livre numérique :

```
<dc:source>http://gallica.bnf.fr/ark:/12148/bpt6k54272182</dc:source>
```

Le type de la ressource est décrit, ainsi que son organisme gestionnaire. :

```
<meta refines="#docnum" property="identifieur-type"  
  scheme="xsd:string">url</meta>
```

```
<meta refines="#docnum" property="meta-auth">www.bnf.fr</meta>
```

6.1.9 Relation

dc:relation contient la notice bibliographique de l'œuvre dans le catalogue de la bibliothèque (cf. balise <reference type="NOTICEBIBLIOGRAPHIQUE"> du refNum).



```
<dc:relation
id="notice">http://catalogue.bnf.fr/ark:/12148/cb30332681g</dc:relation>
<meta refines="#notice" property="identifiant-type" scheme="xsd:string">url</meta>
<meta refines="#notice" property="meta-auth">www.bnf.fr</meta>
```

6.1.10 Coverage

La métadonnée dc:coverage décrit la « couverture » de la publication, en termes spaciaux ou temporels. On y placera donc la date de publication du document d'origine (dans le refNum : <dateEdition>).

Lorsqu'il n'y a pas de <dateEdition> dans les métadonnées (« s.d. » au catalogue, et pas de date trouvée lors de l'exemplarisation et portée au refNum), la métadonnée dc:coverage n'est pas créée.



L'ELEMENT VIDE <DC:COVERAGE/> N'EST PAS ACCEPTE.

6.1.11 Sujet

La métadonnée dc:subject décrit le ou les sujets de la publication.

Le code sujet doit être extrait de l'OAI :

- sujet : champ 680\$d

La Charte d'exemplarisation donne un tableau de correspondance entre codes et valeurs littérales (cf. section 3.3).



Catalogue :
680 6 \$a500\$d500

ePub :
<dc:subject>Science</dc:subject >

6.1.12 Type

La métadonnée dc:type décrit le type (en termes de types de documents : roman, essai, dictionnaire, etc.) de la publication. Les valeurs de cette métadonnée ne sont pas définies par la norme ePub 3.

On y insérera les valeurs suivantes, en fonction de la balise <genre> du refNum :

- <genre>MONOGRAPHIE</genre> : « Monographie »
- <genre>PERIODIQUE</genre> : « Périodique »

6.1.13 Droits

L'élément dc:rights indique les droits d'utilisation des contenus de l'ePub. Sa valeur est fixe :

```
<dc:rights id="#droits">http://gallica.bnf.fr/html/editorial/conditions-dutilisation -  
des-contenus-de-gallica</dc:rights>  
<meta refines="#droits" property="identifier-type" scheme="URI">url</meta>  
<meta refines="#droits" property="meta-auth">www.bnf.fr</meta>
```

6.1.14 Description

On ajoute un élément dc:description pour indiquer la qualité de l'ePub. Sa valeur peut être :

- « gold » : ePub éditorial (tel que produit par les Editions de la BnF, par exemple) ;
- « silver » : ePub patrimonial de qualité supérieure (la production standard des marchés) ;
- « bronze » : ePub patrimonial (l'ePub « aménagé » tel que défini à la section 4.1.2).

6.2 Mentions utiles

6.2.1 Page « À propos de cette édition numérique »

Une page portant les conditions d'utilisation et les mentions utiles est insérée en fin de l'ouvrage.

Cette page comprend une mention de source qui renseigne la bibliothèque dont provient le document, assortie du logo de la BnF lorsque c'est d'elle qu'il s'agit. Cette information est disponible dans le bordereau de traitement aller : Document/bibliographieObjet/refNum:reference type="SOURCE".

Elle comprend également le logo Gallica, et le lien ark vers le document numérisé dans Gallica.

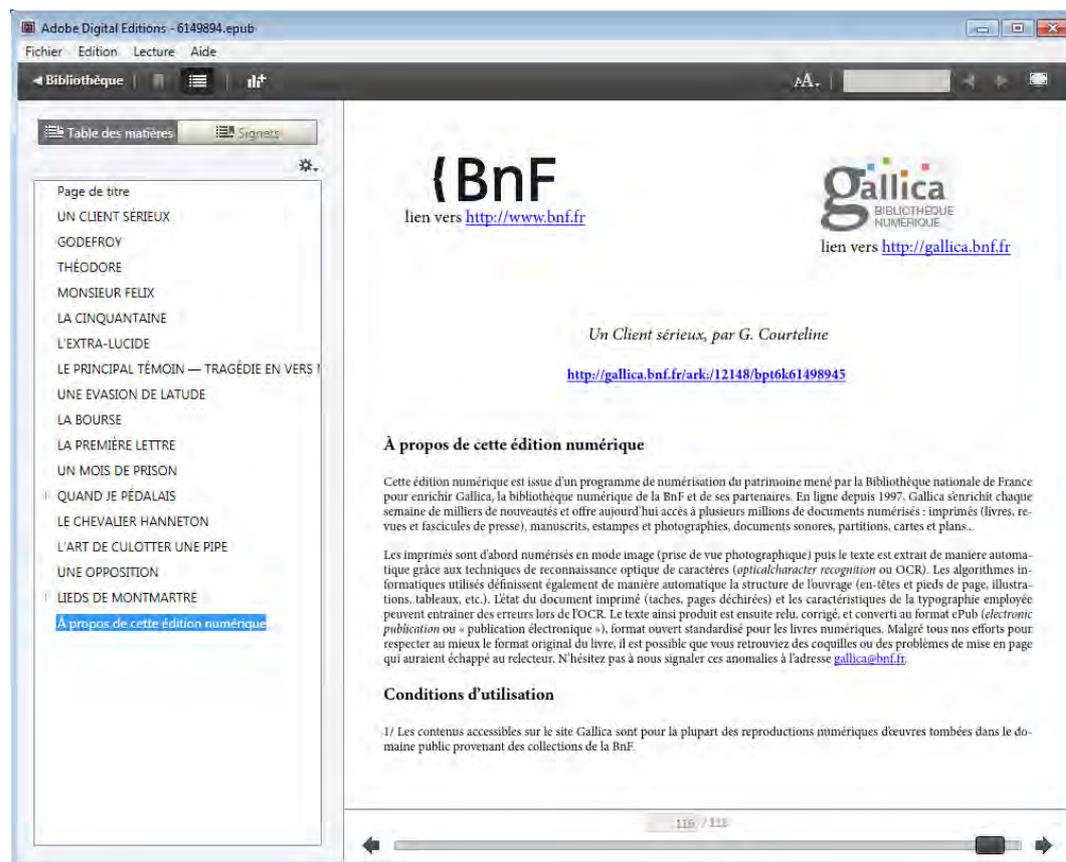
Elle ne porte pas de mention d'ISBN.

Contenu de l'« à propos »

Le contenu est stocké dans un fichier apropos.xhtml (d'identifiant id="apropos") avec deux éléments variables :

- la source du document, au sens de la bibliothèque d'origine, assorti du logo BnF s'il s'agit de la BnF ;
- lien ark vers Gallica dans le lien et dans le texte. Le lien ark est calculé par le prestataire selon la formule idoine.

La mise en forme de cette page est définie par des styles CSS produisant un rendu de ce type :



Page d'à propos

Les images des logos sont des liens hypertextes vers les sites web :

- BnF (sauf dans le cas d'une bibliothèque partenaire : pas de logo)



`bnf.png`

- Gallica



`gallica.png`

Mise en page de l'« à propos »

La mise en page est régie par une feuille de styles CSS spécifique dotée de l'identifiant `id="aproposcss"`.

Place de l'« à propos » dans le livre numérique

L'à propos est inséré en toute fin d'ePub, et une entrée est créée à la fin de la table de navigation de l'ouvrage, avec le texte suivant :

À propos de cette édition numérique



ATTENTION

DANS LE CAS DES EPUB AMENAGES (CF. SECTION 4.1.2), L'A PROPOS EST INSERE APRES LE DERNIER CONTENU DU DOCUMENT, MAIS AVANT LES PAGES DES ZONES TRAITÉES EN MODE IMAGE.

7. CARACTERISTIQUES TECHNIQUES

7.1 Version et profil

La BnF fait le choix du format ePub pour ses livres numériques : il s'agit du seul format standard et libre dans l'industrie du livre numérique, normalisé par une organisation, l'International Digital Publishing Forum (IDPF), qui regroupe les principaux acteurs internationaux du marché.

Néanmoins, ces mêmes acteurs diffusent souvent des lecteurs ePub dotés de fonctionnalités adaptées pour limiter les usages à leurs propres plateformes : la compatibilité des ePub produits n'est donc pas garantie sur toutes les liseuses, tablettes et autres dispositifs de lecture. Pour contourner ce problème, le prestataire veillera à produire des ePub basés sur un socle technique compatible avec la majorité des dispositifs de lecture. Cela implique notamment :

- de ne pas utiliser d'éventuelles extensions propres à un acteur industriel ;
- au sein de la norme ePub, de privilégier une mise en page ne mettant pas en œuvre des capacités avancées (habillage texte/image par exemple).

7.1.1 Livre numérique et accessibilité

Le format ePub 3 présente l'avantage de réunir les avantages d'un format ouvert, grand public et interopérable, tout en étant doté des mécanismes garantissant un meilleur accès aux contenus numériques. Comparé à Daisy, il offre l'avantage d'offrir une meilleure accessibilité aux personnes souffrant d'une déficience visuelle légère ou moyenne, personnes qui ne se seraient pas forcément tournées vers le format Daisy, qui est associé au handicap visuel lourd. De manière générale, une meilleure accessibilité profite à tous, à tout moment, et c'est bien l'esprit qui a conduit le consortium Daisy à rejoindre le standard ePub 3 en le faisant évoluer sur cet aspect.

Le format ePub 3 sera donc utilisé pour la production, mais en veillant à favoriser la compatibilité descendante :

- produire les deux formats de table de navigation (.ncx et .html, cf. section 5.5) ;
- conserver les métadonnées ePub 2, notamment la spécification de la couverture :

```
<meta name="cover" content="cover-image"/>
```

- ne pas placer de références croisées sur des balises HTML5 ;
- etc.



Lors de la phase de test, si une version mineure de la norme ePub 3 est disponible (ePub 3.1 ou 3.2 par exemple), le prestataire et la BnF évalueront sa stabilité et sa maturité avant application éventuelle.

De plus, le prestataire veillera à ne pas utiliser les éléments à risques (tant en termes de compatibilité descendante que de compatibilité ePub 3) introduits par ePub 3 :

- contenus multimédia,
- scripting,
- synchronisation texte-son.



Un risque demeure quant au balisage HTML 5 des contenus, pour lequel une évaluation du comportement des liseuses ePub 2 devra être menée en phase de test.

Format pivot pour l'accessibilité

Afin de favoriser l'accessibilité des contenus numériques de la bibliothèque numérique Gallica, il est demandé au prestataire de fournir les contenus des ePub produits selon un format pivot XML conforme à norme DAISY 3 (Z39.86-2002 et 2005) via la DTD DAISY XML (ou *Daisy Digital Talking Book*, DTBook v2005-3).



Ce format pivot sera notamment mis à disposition des associations au service des aveugles et des malvoyants, à leur demande, afin de faciliter la conversion vers tout format de diffusion adapté à leur mission.

Le prestataire devra donc mettre en place un mapping de son propre format pivot vers le format pivot DAISY XML. Ce mapping sera évalué par la BnF en phase de test puis appliqué automatiquement par le prestataire en phase de production. L'objectif est de transcrire dans le format pivot DAISY XML l'intégralité des efforts de structuration logique et d'annotation sémantique appliqués aux contenus de l'ePub (cf. section 5.2).

Le format pivot sera délivré à la BnF selon les attendus de la norme DAISY :

- les contenus textuels sous la forme d'un fichier XML, conforme à la DTD <http://www.daisy.org/z3986/2005/dtbook-2005-3.dtd>,
- les éventuels graphiques, illustrations, cartes et autres éléments image référencés dans les contenus textuels, sous la forme de fichiers image.

7.1.2 Variante *fixed layout*

La cible principale des conversions d'ouvrage en mode livre électronique étant des œuvres où le texte domine, l'usage de l'ePub standard (mode « flot de texte recomposable ») couvre la plupart des besoins. Cependant, le prestataire présentera son aptitude à traiter des genres où la mise en page prime :

- livres jeunesse,
- recueils de planche illustrée, d'estampes,
- etc.

Pour ces genres, la mise en œuvre de la variante *fixed layout* d'ePub 3 (<http://www.idpf.org/epub/fxl>) sera envisagée conjointement par le prestataire et la BnF, notamment selon ces aspects :

- adaptation éditoriale des contenus aux caractéristiques et limitations du format *fixed layout* ;

- utilisabilité du format *fixed layout* sur le parc des dispositifs de lecture.

7.2 Feuilles de styles CSS

L'ePub peut comporter une ou plusieurs feuilles de styles CSS. Elles doivent être regroupées dans un répertoire nommé « Styles ».



La BnF et le prestataire étudieront la possibilité de décliner les feuilles de styles en fonction des genres documentaires à traiter :

- monographie/périodique,
- roman/essai
- etc.

7.3 Découpage des contenus

L'ePub peut comporter une ou plusieurs fichiers de contenu texte, selon le volume du document d'origine. Le prestataire veillera à ce que les fichiers élémentaires conservent une taille raisonnable (quelques centaines de ko), afin d'être lus par des liseuses à encre électronique d'ancienne génération.

De même, il est conseillé qu'à chaque fichier de contenu texte, ne soit lié pas plus de 10 Mo de fichiers de contenu image.

7.4 Polices de caractères

Le format d'ePub permet d'incorporer des polices de caractères, notamment pour contourner les limitations des liseuses à reproduire la totalité du jeu de caractères Unicode. Selon les besoins, le prestataire pourra donc incorporer une ou plusieurs polices de caractères, mais il devra s'assurer qu'elles sont bien libres de droits pour une telle utilisation.

L'ePub peut comporter une ou plusieurs polices. Elles doivent être regroupées dans un répertoire nommé « Polices ».

7.5 Images

Les images à insérer dans le texte devront être redimensionnées de manière à :

- réduire la taille globale des images afin de faciliter leur manipulation par les liseuses et tablettes ;



La taille maximale d'une image est fixée à 2 millions de pixels.

- garder leur netteté et leur lisibilité, notamment lorsqu'elles comprennent du texte ;
- conserver les tailles relatives des images de l'original. Ainsi, une image occupant la moitié de la justification dans l'ouvrage original devra

occuper 50 % de la largeur de l'ePub une fois celui-ci affiché dans un logiciel de lecture d'ePub ;

- permettre le zoom sur les images, sur les liseuses et logiciels de lecture qui disposent de cette fonction, sans pour autant pénaliser les autres.

Ces exigences ne seront généralement atteintes que par un compromis.

Le format de stockage des images est laissé à l'appréciation du prestataire, sachant que le format JPEG offre le meilleur ratio qualité/compression.



Le nom des images doit reprendre l'identifiant de la zone ALTO qui leur correspond, afin de permettre le contrôle de leur typage à terme. Exemple de balisage d'une illustration dans le fichier Alto :

```
<Illustration ID="PAG_00000001_IL000001" HEIGHT="999" WIDTH="2614"  
HPOS="36" VPOS="322" />
```

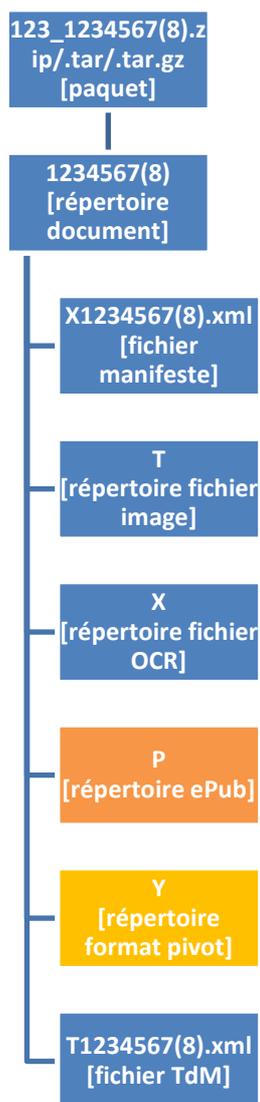
Dans cet exemple, il faut reprendre comme nom du fichier image PAG_00000001_IL000001, où PAG_NNNNN est l'identifiant de la page, et ILNNNNN est l'identifiant de l'illustration.

Toutes les images doivent être regroupées dans un répertoire nommé « Images ».

7.6 Livraison

La livraison des fichiers ePub doit être réalisée conformément aux directives générales du Référentiel de livraison de document numérique.

Le paquet d'un document numérique incluant un ePub et/ou un format pivot accessible doit présenter l'organisation suivante :



Livraison

Le répertoire P doit contenir le fichier de l'ePub :

- corps : l'identifiant du document numérique, sur 7 chiffres
- extension : .epub

Le répertoire Y doit contenir le format pivot dédié à l'accessibilité, sous la forme d'une archive empaquetant les fichiers du format pivot :

- corps : l'identifiant du document numérique, sur 7 chiffres
- extension : .zip

L'archive du format pivot dédié à l'accessibilité doit contenir :

- le format pivot des contenus textuels :
 - corps : l'identifiant du document numérique, sur 7 chiffres
 - extension : .xml
- les éventuels graphiques, illustrations, etc. référencés dans les contenus textuels, sous la forme d'un fichier image par élément référencé.

- corps : le nommage décrit à la section 7.5 doit être appliqué
- extension : .tif ou .jp2

7.6.1 Formats

ePub

L'ePub doit être conforme au format ePub 3 (ou une version mineure ultérieure, cf. section 7.1.1).

Format pivot accessible

Le format pivot des contenus textuels doit être conforme à la DTD DAISY XML (DTBook 2005-3, cf. section 7.1.1) ou à un autre format équivalent demandé par la BnF.

Le format des éventuelles images référencées dans les contenus textuels doit être celui utilisé pour les images de conservation (TIFF ou JPEG2000, selon les marchés). Leur résolution sera également celle des images de conservation.

8. REJET DES DOCUMENTS A TRAITER EN EPUB

Le prestataire mettra en place un processus de déqualification des documents non éligibles ePub en considérant les points suivants :

- critères bibliographiques (ex : monographies de type presse, cf. section 4.2) ;
- déqualification à cause de contraintes techniques (ex : courbure inévitable, cf. section 4.1.2) ou de problèmes liés à la qualité de l'original.



Les ePub sont produits à l'UD. Lorsque certaines UD ne sont pas éligibles ePub, l'UC est produite avec un rapport détaillant les causes de rejets des UD non produites.

La liste de causes de rejets est non exhaustive. Cette liste sera complétée pendant les phases de test et de production dans une dynamique d'amélioration continue.

Ce processus vise à identifier au plus tôt les ouvrages qui ne pourront pas être produits en ePub afin de limiter le nombre d'opérations de production inutiles si la déqualification est confirmée.

8.1 Cas des ePub « aménagés »

Comme il a été dit précédemment (section 4.1.2), les documents non éligibles à une production ePub HQ peuvent néanmoins être produits en ePub aménagés, sous réserve qu'ils valident certains critères, par exemple qu'ils présentent au maximum x % de pages impactées par des défauts (blocs illisibles, courbures, etc.). Dans le cas contraire, ils seront définitivement rejetés.

Le prestataire fournira en phase de test un catalogue ayant pour objectif de lister les catégories d'unités documentaires, sur la base d'exemples, pour lesquelles il souhaite l'accord de la BnF pour un rejet. Ce document pourra être mis à jour au fil du marché, en fonction des nouvelles typologies documentaires éventuellement rencontrées.

Le prestataire et la BnF se mettront également d'accord sur les critères d'éligibilité à la production sous forme d'ePub aménagés.

9. CONTROLES APPLIQUES PAR LA BNF

9.1 Contrôles automatiques

Les contrôles automatiques exposés ci-après sont appliqués à tous les ePub livrés à la BnF.

9.1.1 Contrôle technique

L'ePub doit passer avec succès le contrôle appliqué par l'outil EpubCheck. Aucun retour de l'outil Epubcheck de type ERROR ou WARN ne sera toléré.



La version ePubCheck utilisée sera celle disponible au lancement du marché. Si une nouvelle version apparaît en cours de marché, le prestataire et la BnF s'accorderont pour l'utiliser ou non.

9.1.2 Contrôle de l'ark

On contrôle que l'ark du document numérique généré par le prestataire est conforme à la méthode de calcul communiquée par la BnF.

Cet ark doit être présent sur la page d'à propos, dans la balise <metadata> du fichier OPF et dans l'attribut content de la balise <head><meta> de la table de navigation .ncx.



L'ePub doit valider ces conditions pour être accepté en chaîne d'entrée.

9.1.3 Contrôle du format général de l'ePub



L'ePub doit valider toutes les conditions énoncées ci-après pour être accepté.

Le répertoire principal de l'archive doit s'appeler « OEBPS ».

Manifeste et table de navigation

Il doit y avoir un seul fichier manifeste OPF.

Il doit exister un seul fichier de table de navigation « toc.ncx ». Il doit exister un seul fichier de table de navigation « toc.xhtml ».

Texte

Les fichiers de contenu HTML/XHTML et les fichiers de structure (.opf et .ncx) doivent être encodés en UTF-8. Tout fichier de contenu HTML/XHTML doit être stocké dans le répertoire principal OEBPS.

Styles

Au moins une feuille de style CSS doit être présente. Tout fichier CSS doit être stocké dans un sous-répertoire nommé « Styles » (dans le répertoire principal « OEBPS », soit OEBPS/Styles).

Polices de caractères incorporées

L'inclusion de polices dans l'ePub est optionnelle. Toute police doit être stockée dans un sous-répertoire nommé « Polices » (dans le répertoire principal « OEBPS », soit OEBPS/Polices).

Images

Toute image doit être stockée dans un sous-répertoire nommé « Images » (dans le répertoire principal « OEBPS », soit OEBPS/Images).

Les logos BnF et Gallica doivent être identiques à ceux fournis par la BnF.

Le niveau de qualité, maintenu lors de la compression par rapport à l'image originale, doit être supérieur ou égal à un minimum défini, qui est de 80 %. Ce taux de qualité est vérifié par la BnF à l'aide de l'outil ImageMagick (<http://www.imagemagick.org>) :

```
>identify -verbose fichier_image
...
Compression: JPEG
Quality: 80
...
```

9.1.4 Contrôle de structure général de l'Epub



L'ePub doit valider toutes les conditions énoncées ci-après pour être accepté, sauf mention contraire (avertissement WARN).

Contrôle du fichier OPF

Le fichier de structure .opf doivent être encodés en UTF-8.

Contrôles pour la balise <metadata>

La balise <metadata> doit être présente une seule fois.



Pour assurer la compatibilité descendante avec les liseuses ePub 2, elle doit contenir la ligne <meta name="cover" content="cover-image"/>.

Le contenu des balises suivantes est contrôlé, avec un retour de contrôle ERROR ou WARN selon les cas :

- dc:format à "application/epub+zip" – ERROR
- dc:identifiant id="dcidid" à l'ark du document concaténé avec son extension « .epub » (cf. section 6.1.1) – ERROR
- dc:title au titre du document, selon les variantes décrites à la section 6.1.2 – WARN
- dc:creator au(x) auteur(s) du document, selon les variantes décrites à la section 6.1.3 – ERROR
- dc:contributor au(x) contributeur(s) du document, selon les variantes décrites à la section 6.1.4 – ERROR
- dc:publisher à l'éditeur du document (cf. section 6.1.5) – ERROR
- dc:date à une date récente (la date de réalisation de l'ePub) (cf. section 6.1.6) ;
- dcterms:issued à la date de publication de l'œuvre originelle (cf. section 6.1.6) – WARN
- dc:language à la langue principale de l'œuvre (cf. section 6.1.7) – ERROR
- dc:source à l'URL Gallica et à l'ark du document (cf. section 6.1.8) – ERROR
- dc:relation à l'URL catalogue du document et l'ark de la notice du document (cf. section 6.1.9) – ERROR
- dc:coverage à la date de publication de l'œuvre originelle (cf. section 6.1.6) – ERROR
- dc:subject à la cote bibliographique du document (cf. section 6.1.11) – ERROR
- dc:type au genre du document (cf. section 6.1.12) – ERROR
- dc:rights : aux conditions d'utilisation dans Gallica (cf. section 6.1.13) – ERROR
- dc:description : à la qualité éditoriale de l'ePub, soit « bronze » ou « silver » pour un marché de numérisation de masse, et « gold » pour un ePub éditorial (cf. section 6.1.14) – ERROR

Contrôles pour la balise <manifest>

La balise <manifest> doit être présente une seule fois et comprendre obligatoirement les items suivants :

- id="apropos", href="apropos.xhtml", media-type="application/xhtml+xml"
- id="aproposcscs", media-type="text/css"
- id="coverhtml", media-type="application/xhtml+xml"
- id="cover-image", media-type="image/jpeg", properties="cover-image"
- id="ncx", href="toc.ncx", media-type="application/x-dtbncx+xml"
- id="toc", href="toc.xhtml", media-type="application/xhtml+xml", properties="nav"
- id="logogallica", href="logo_gallica.png", media-type="image/png"
- id="logobnf", href="bnf.png", media-type="image/png"

Contrôles pour la balise <spine>

La balise <spine toc="ncx"> doit être présente une seule fois et doit comprendre les itemref suivants (ainsi bien sûr que les références vers les autres contenus de l'ePub) :

- idref="apropos"
- idref="coverhtml"

```
<spine toc="ncx">
  <itemref idref="coverhtml" />
  ...
  <itemref idref="apropos" />
</spine>
```

Contrôle du fichier toc.xhtml

Table de navigation

Le fichier .xhtml doit être encodé en UTF-8.

Le titre présent dans la balise <title> doit être identique au contenu de la balise dc:title du fichier OPF.

La balise <nav epub:type="toc"> doit être présente une seule fois. Elle doit être suivie par la balise suivante :

```
<h1>Table des matières</h1>
```

L'entrée concernant la page de titre doit être conforme à l'intitulé demandé par la BnF (cf. section 5.4.2). Le nom du fichier .xhtml n'est pas imposé :

```
<li id="navpoint-2">
  <a href="xxxx.xhtml">Page de titre</a>
</li>
```

L'entrée concernant le message d'« à propos » doit être conforme à l'intitulé demandé par la BnF (cf. section 6.2.1). La valeur de l'attribut href de la balise <content> doit correspondre à « apropos.xhtml » :

```
<li id="navpoint-32">
  <a href="apropos.xhtml">À propos de cette édition électronique</a>
</li>
```

Table des repères

La balise <nav epub:type="landmarks"> (cf section 5.2.2) doit être présente une seule fois. Elle doit être suivie par la balise suivante :

```
<h1>Repères</h1>
```

Elle doit être structurée sur un seul niveau hiérarchique.

Elle doit inclure obligatoirement les références suivantes :

- <a epub:type="toc"> : cette entrée référence la table des matières de l'ePub (au format HTML) ;

- `<a epub:type="bodymatter">` : cette entrée référence le début du contenu principal de l'ePub.

```
<ol>
  <li><a epub:type="toc" href="#toc">Table des matières</a></li>
  <li><a epub:type="bodymatter" href="ebpt6k6289050k_c01.xhtml">Début
    du texte</a></li>
</ol>
```

Table des numéros de pages physiques (folios)

La balise `<nav epub:type="page-list">` (cf section 5.2.2) doit être présente une seule fois. Elle doit être suivie par la balise suivante :

```
<h1>Pages</h1>
```

La table doit obligatoirement inclure une référence à toutes les pages numérotées du document d'origine :

```
<nav epub:type="page-list">
<h1>Pages</h1>
<ol>
  <li><a href="pagedetitre.xhtml#pageii">II</a></li>
  ...
  <li><a href="preface.xhtml#pagev">V</a></li>
  <li><a href="preface.xhtml#pagevi">VI</a></li>
  <li><a href="chapitre001.xhtml#page001">1</a></li>
  <li><a href="chapitre001.xhtml#page002">2</a></li>
  ...
  <li><a href="chapitre001.xhtml#page256">256</a></li>
</ol>
```

Pour chaque entrée de la table, le lien doit pointer vers une ancre placée dans le contenu numérique, ancre qui référence un numéro de page physique du document d'origine.

Contrôles pour la balise `<guide>` (compatibilité ePub 2)

La balise `<guide>` doit être présente une seule fois et doit inclure :

- `type="cover"` : cette entrée référence la page de couverture de l'ePub (au format HTML).

```
<guide>
  <reference href="Text/couv.xhtml" title="Page de couverture"
    type="cover" />
</guide>
```

Contrôle du fichier `toc.ncx` (compatibilité ePub2)

Le fichier `.ncx` doit être encodé en UTF-8.

Le titre présent dans `<docTitle><text>` doit être identique au contenu de la balise `dc:title` du fichier OPF.

La valeur de l'attribut `content` de la balise `<head><meta>` doit être identique à l'identifiant de l'ePub (`dc:identifier`), soit en fait l'ark de l'ePub (cf. section 6.1.1).

```
<meta content="ark:/12148/bpt6k54272182/w1.epub" name="dtb:uid" />
```

Les informations présentes sur les attributs id et playorder de la dernière balise <navpoint> doivent être cohérentes, et correspondre au nombre de balises <navpoint> présentes dans le fichier.

L'entrée concernant la page de titre doit être conforme à l'intitulé demandé par la BnF (cf. section 5.4.2). Le nom du fichier .xhtml n'est pas imposé :

```
<navPoint id="toc2" playOrder="2">
  <navLabel>
    <text>Page de titre</text>
  </navLabel>
  <content src="xxx.html"/>
</navPoint>
```

L'entrée concernant le message d'« à propos » doit être conforme au message demandé par la BnF (cf. section 6.2.1). La valeur de l'attribut src de la balise <content> doit correspondre à « apropos.xhtml » :

```
<navPoint id="navpoint-32" playOrder="32"> <!-- remplacer 32 par le dernier
                                             numéro + 1 -->
  <navLabel>
    <text>À propos de cette édition numérique</text>
  </navLabel>
  <content src="apropos.xhtml"/>
</navPoint>
```

9.2 Contrôles visuels

Un contrôle par échantillonnage visuel opère sur des lots d'ePub constitués selon un plan d'échantillonnage. Ce plan permet de définir un échantillon en termes de documents ePub contrôlés et d'extraits de contenu contrôlés au sein des documents.



Les modalités de ce contrôle seront détaillées dans une charte de contrôle ePub élaborée conjointement par la BnF et le prestataire.

Les éléments suivants sont contrôlés visuellement (liste non exhaustive) :

- page de couverture : choix correct de la page de titre, netteté de l'image, mise en forme de l'image
- pages liminaires: choix de la page de titre, restitution en mode texte de la page de titre, présence et restitution des autres pages liminaires
- table de navigation : vérification de la table (niveaux de titre, présence des éléments obligatoires, coquilles, etc.)
- page d'à propos : contenu, mise en forme
- styles de paragraphe : vérification de la mise en forme des différents niveaux de titre, des styles d'alignement de paragraphe, etc.
- titres courants, pieds de page : vérification que d'éventuels titres courants n'ont pas été introduits dans le corps de l'ouvrage
- typographie : respect du code typographique, gestion des césures, etc.

- lettrines : vérification du traitement correct des lettrines
- illustrations : vérification de la mise en page des illustrations (taille, positionnement, etc.)
- liens et références croisées : vérification de leur fonctionnement
- qualité éditoriale : vérification de la montée en qualité des contenus texte
- cohérence du contenu : contenu présent dans le document source et absent de l'ePub, contenu étranger introduit dans l'ePub
- cohérence de l'ordre de lecture.

Ces contrôles génèrent des défauts qualifiés de « majeurs » ou de « mineurs », comme détaillé dans les tableaux suivants. Le rejet de l'ePub contrôlé sera prononcé en fonction du nombre de défauts constatés pour chaque catégorie de défauts (majeurs ou mineurs).

9.2.1 Défauts majeurs

Qualité éditoriale

Le rejet de l'ePub est prononcé si le taux qualité mesurant le taux des erreurs majeurs résiduelles dans le texte contenu dans l'ePub n'atteint pas la valeur attendue. Les taux sont mesurés selon la définition donnée section 5.1 et la valeur du taux est donnée section 5.1.4.

Autres défauts majeurs

La présence d'un défaut majeur parmi les catégories du tableau suivant entraîne le rejet de l'ePub contrôlé.

Types de défaut majeur	Détails
Couverture	<p>La couverture de l'ePub souffre d'un défaut majeur :</p> <ul style="list-style-type: none"> • couverture absente, mauvais choix de la page dans le document d'origine • image mal recadrée ou mal redressée • image de qualité insuffisante (dimensions, compression, profondeur de couleurs) • mauvaise mise en page de la couverture
Pages liminaires	<p>La page de titre est absente ou n'est pas correcte :</p> <ul style="list-style-type: none"> • elle n'est pas restituée en mode texte • elle contient au moins une coquille • elle est suivie ou précédée de page(s) de faux-titre, qui devraient être supprimées <p>Les autres pages de début de l'ouvrage sont absentes ou incorrecte :</p> <ul style="list-style-type: none"> • dédicace, remerciements, etc.
Table de navigation	<p>Au moins un défaut majeur dans la table de navigation de l'ePub :</p> <ul style="list-style-type: none"> • titre de la table erroné • oubli d'une entrée obligatoire (« Page de titre », « A propos de cette édition numérique », « Notes »)

	<ul style="list-style-type: none"> • oubli d'une entrée • mauvaise hiérarchisation des entrées • une entrée ne pointe pas vers le bon emplacement dans le contenu • au moins une coquille dans une entrée • numérotation des entrées présente dans le document d'origine et absente ou incorrecte dans la table
Table des repères	<p>Au moins un défaut majeur dans la table des repères de l'ePub :</p> <ul style="list-style-type: none"> • table absente • titre de la table erroné • entrée(s) obligatoire(s) manquante(s) • entrée manquante ou entrée sans objet • une entrée ne pointe pas vers le bon emplacement dans le contenu • au moins une coquille dans une entrée • structure de la table erronée
Table des folios	<p>Au moins un défaut majeur dans la table des folios de l'ePub :</p> <ul style="list-style-type: none"> • table absente • titre de la table erroné • entrée manquante ou entrée sans objet • une entrée ne pointe pas vers le bon emplacement dans le contenu • au moins une coquille dans une entrée • structure de la table erronée
Page à propos	Le contenu de la page « A propos de cette édition numérique » est erroné
Liens	<p>Tous les liens sont erronés ou absents :</p> <ul style="list-style-type: none"> • lien entre appel de note et corps de note • lien entre le numéro de figure ou table et la figure ou la table • lien vers image du mode « ePub aménagé »
Ordre de lecture	<p>Non-respect de l'ordre de lecture du document source, désorganisation du contenu :</p> <ul style="list-style-type: none"> • au niveau des pages • au niveau du document
Élément(s) étranger(s) à l'ePub	L'ePub contient un contenu qui n'est pas présent dans le document source.
Élément(s) manquant(s) dans l'ePub	<p>Il manque dans l'ePub un contenu qui est présent dans le document source :</p> <ul style="list-style-type: none"> • mot, ligne, paragraphe, page • image, tableau • ...
Élément(s) inutile(s) dans l'ePub	<p>L'ePub contient un contenu qui ne devrait pas être présent :</p> <ul style="list-style-type: none"> • pages de plats, de faux titre • table de navigation doublée dans le contenu, ou table de type sommaire avec liens hypertextes • éléments graphiques (cul-de-lampe, tiret, tampon)

	<ul style="list-style-type: none"> • annotations manuscrites
Structuration logique des contenus	<p>Au moins un type de défaut récurrent de structuration logique est constaté :</p> <ul style="list-style-type: none"> • aucune structuration logique n'est présente • les règles de structuration logique (section 5.2.2) sont systématiquement mal appliquées
Caractères	<p>Un défaut systématique affecte l'encodage des caractères et rend leur lecture impossible :</p> <ul style="list-style-type: none"> • des caractères sont mal restitués • des caractères spéciaux du document d'origine sont mal restitués
Typographie	<p>Au moins un type de défaut typographique récurrent affecte la présentation de la totalité du contenu texte :</p> <ul style="list-style-type: none"> • code typographique français non appliqué : absence d'espace insécable, espaces doublées, etc. • enrichissement typographique absent ou erroné : italique, gras, exposant, etc.
Mise en page	<p>Au moins un type de défaut de mise en page récurrent affecte la présentation de la totalité du contenu :</p> <ul style="list-style-type: none"> • feuille de styles non respectée : titres, légendes, listes, notes de bas de page, etc. • mise en page incorrecte ou non homogène : marges, alignements, etc. • numérotation erronée de titres, listes numérotées, notes, etc. • lettre ornée non traitée • traitement des tableaux • etc.
Mauvaise restitution des images numérisées	<p>Toutes les images présentes dans l'ePub présentent un défaut, parmi :</p> <ul style="list-style-type: none"> • Lisibilité de l'image numérisée insuffisante par rapport à la qualité du support original • Image couleur de l'original traitée en NDG ou N&B • Image NDG de l'original traitée en N&B • Défaut manifeste de pixellisation • Défaut d'orientation des images par rapport à l'original • La taille relative de l'image affichée dans l'ePub par rapport à celle de la page d'origine n'est pas respectée • La taille effective de l'image (sa définition, en pixels) est insuffisante et ne permet pas le zoom
Autre défaut majeur - commentaire libre	<p>Tout défaut qui peut être considéré comme majeur et qui n'aurait pas été envisagée dans les cas précédents. Dans ce cas, la BnF soumettra ces nouvelles règles au titulaire et validera la date de mise en œuvre de leur application.</p>

9.2.2 Défauts mineurs

Qualité éditoriale

Le rejet de l'ePub est prononcé si le taux qualité mesurant le taux des erreurs mineures résiduelles dans le texte contenu dans l'ePub n'atteint pas la valeur attendue. Les taux sont mesurés selon la définition donnée section 5.1 et la valeur du taux est donnée section 5.1.4.

Autres défauts mineurs

La présence de défauts mineurs entraîne le rejet de l'ePub contrôlé si le taux des défauts mineurs est supérieur à un seuil.

Le taux des défauts mineurs est calculé avec la formule suivante :

$$\text{cardinal des défauts mineurs dans l'échantillon de contrôle} \div \text{cardinal des mots dans l'échantillon de contrôle}$$

Types de défaut mineur	Détails
Segmentation	Un défaut affectant la segmentation du texte courant : <ul style="list-style-type: none">• un paragraphe du document source est scindé en deux paragraphes dans l'ePub• deux paragraphes du document source sont accolés dans l'ePub• un retour à la ligne intempestif dans l'ePub
Typographie	Un défaut tel que ceux décrits dans le tableau des défauts majeurs
Mise en page	Un défaut tel que ceux décrits dans le tableau des défauts majeurs
Structuration logique	Un défaut tel que ceux décrits dans le tableau des défauts majeurs
Mauvaise restitution des images numérisées	Un défaut tel que ceux décrits dans le tableau des défauts majeurs
Liens	Un défaut tel que ceux décrits dans le tableau des défauts majeurs
Pages spéciales	Défaut sur une page spéciale du document original : catalogue d'éditeur, publicité, errata, etc.
Autre défaut mineur - commentaire libre	Tout défaut qui peut être considéré comme mineur et qui n'aurait pas été envisagée dans les cas précédents. Dans ce cas, la BnF soumettra ces nouvelles règles au titulaire et validera la date de mise en œuvre de leur application.

Les valeurs attendues pour ce taux qualité sont définies dans le tableau suivant.

Typologie	Taux attendu
Qualité éditoriale « argent »	Défauts mineurs < 0,1 % (taux qualité : 99,9 %)
Qualité éditoriale « bronze »	Défauts mineurs < 0,125 % (taux qualité : 99,875 %)

10. TERMINOLOGIE

10.1 Glossaire

ALTO : ce format est couramment utilisé pour la conversion de texte à partir d'images. Il est basé sur XML et géré par un schéma XML. Il conserve toutes les coordonnées du contenu (texte, illustrations, graphiques) présent dans l'image et permet la superposition de l'image et du texte (fichier PDF multicouche) ainsi que la surbrillance des mots recherchés lors d'une requête.

Voir aussi :

http://www.bnf.fr/fr/professionnels/num_conversion_texte/s.num_conversion_texte_ocr.html

et <http://www.loc.gov/standards/alto>

ark : ARK (*Archival Resource Key*) est un système d'identifiants mis en place par la California Digital Library (CDL), et qui a vocation à identifier des objets de manière pérenne. Il peut s'agir d'objets de tous types, physiques (table, livre), numériques (livre numérisé...) ou même immatériels (concepts, ...).

Voir aussi :

http://www.bnf.fr/fr/professionnels/s_informer_autres_numeros/a.ark_autres_numeros.html

Daisy : DAISY (*Digital Accessible Information System*) est une norme pour livres numériques proposée par le consortium DAISY (<http://www.daisy.org>). Ces livres sont destinés aux personnes empêchées de lire (aveugles, malvoyants, dyslexiques, handicapés voyants mais ayant du mal à tourner les pages d'un livre, etc.). Il s'agit de livres numériques structurés, avec contenu audio ou non, permettant une navigation aisée à l'intérieur du contenu.

epubcheck : EpubCheck est un outil de validation des fichiers EPUB, version 2.0 et ultérieures. Il peut détecter de nombreux types d'erreurs : structure du manifeste OPF, cohérence interne de la structure, validation des contenus XHTML, etc.

Voir aussi validator.idpf.org et code.google.com/p/epubcheck

Fixed layout : la variante *fixed layout* du standard ePub permet de produire des livres numériques dotés d'une mise en page fixe. Le contenu n'est alors plus recomposé à la volée en fonction de la taille d'écran, de l'orientation de l'écran, etc. Cette variante est adaptée aux ouvrages à forte composante graphique (livres jeunesse, manuels scolaires, beaux-livres, etc.).

Qualité ePub : la qualité liée au processus utilisé pour produire un epub donné : « gold » : ePub éditorial (tel que produit par les Editions de la BnF, par exemple) ; « silver » : ePub patrimonial de qualité supérieure (la production standard des marchés) ; « bronze » : ePub patrimonial (l'ePub « aménagé » tel que défini à la section 4.1.2)

Qualité éditoriale : une qualité de type « éditoriale » caractérise des contenus présentant un faible taux d'erreurs. Ce taux peut varier de 99,96 % à 99,99 % (une coquille tous les 10 000 caractères), selon le niveau d'exigence de l'éditeur et la nature du processus de production des contenus.

Qualité OCR : lors d'un processus d'océrisation, la qualité finale du texte produit est variable, et dépend notamment de la nature du document originel. Cette qualité est généralement représentée par un taux qualité exprimé en %.

refNum : pour gérer les métadonnées de production des documents numériques, la BnF a mis au point le schéma XML refNum, schéma propre à l'établissement et

utilisé par les applications de contrôle des données livrées par les prestataires et les ateliers. La structure du schéma refNum est proche du standard METS, mais simplifiée pour faciliter la production.

Voir aussi :

http://www.bnf.fr/fr/professionnels/num_metadonnees/s.num_metadonnees_documents.html et <http://bibnum.bnf.fr/refNum/documentation/refNum.html>

10.2 Abréviations

ALTO : *Analyzed Layout and Text Object*

ark : *Archival Resource Key*

Daisy : *Digital Accessible Information SYstem*

ePub : *Electronic Publication*

XML : *Extensible Markup Language*