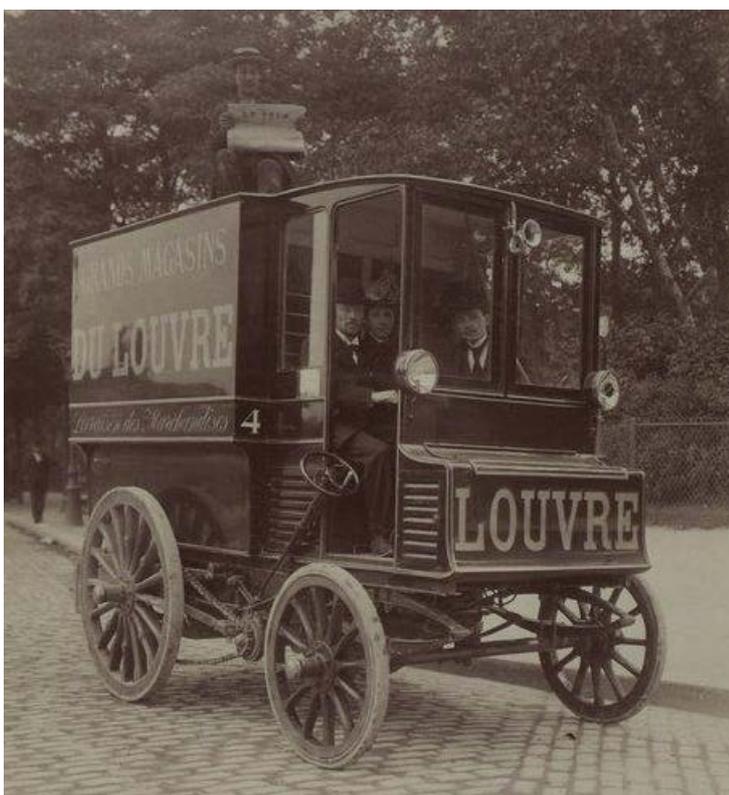


Référentiel de livraison de document numérique



<http://gallica.bnf.fr/ark:/12148/btv1b84333337/f66.item>

TABLE DES MATIERES

1.	INTRODUCTION	3
2.	DOCUMENTS APPLICABLES ET DE REFERENCE	4
3.	DESCRIPTION DU SERVICE DE LIVRAISON	5
4.	DESCRIPTION GENERALE D'UN DOCUMENT NUMERIQUE	6
5.	ORGANISATION DES PAQUETS DE LIVRAISON	7
<hr/>		
5.1	ORGANISATION DES LIVRAISONS - MARCHES REFNUM	8
5.2	ORGANISATION DES LIVRAISONS - MARCHES METS	9
6.	DESCRIPTION DES TYPES DE LIVRAISON	10
<hr/>		
6.1	LIVRAISON INITIALE	10
6.2	LIVRAISON COMPLEMENTAIRE	10

1. INTRODUCTION

Ce référentiel s'inscrit dans les opérations de numérisation réalisées pour le compte de la BnF.

Le présent référentiel définit les consignes à respecter pour la livraison des documents numériques à la Bibliothèque nationale de France.

Les livraisons concernent deux types de documents numériques :

- les documents numériques directement créés à partir d'un objet physique : dans ce cas il convient de se référer au chapitre « livraison initiale ».
- les documents numériques qui sont complétés par des éléments supplémentaires (OCR, table des matières, livre numérique, format pivot adaptable) : dans ce cas il convient de se référer au chapitre « livraison complémentaire ».

De manière générale, les livraisons s'effectuent par dépôt sur un serveur FTP.

2. DOCUMENTS APPLICABLES ET DE REFERENCE

Les référentiels de numérisation de la BnF sont disponibles à cette adresse :
http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.numerisation_referentiels_bnf.html

Intitulé	Description	Référence
<i>Référentiel d'enrichissement des métadonnées</i>		http://www.bnf.fr/documents/ref_num_metadonnees_refnum3.pdf
<i>Référentiel d'enrichissement des métadonnées - version METS</i>		http://www.bnf.fr/documents/ref_num_metadonnees_mets.pdf
<i>Référentiel de numérisation des documents opaques</i>		http://www.bnf.fr/documents/ref_num_doc_opaques.pdf
<i>Référentiel de numérisation des documents transparents</i>		http://www.bnf.fr/documents/ref_num_doc_transparents.pdf
<i>Référentiel OCR</i>		http://www.bnf.fr/documents/ref_num_ocr_v2.pdf
<i>Référentiel traitement des tables</i>		http://www.bnf.fr/documents/ref_num_traitement_tables.pdf
<i>Référentiel EPUB 3</i>		http://www.bnf.fr/documents/ref_num_epub3.pdf
<i>Référentiel DAISY</i>		http://www.bnf.fr/documents/ref_num_daisy.pdf

3. DESCRIPTION DU SERVICE DE LIVRAISON

Les documents numérisés doivent être livrés par réseau via un serveur FTP. Le protocole utilisé doit être conforme aux RFC959¹ et RFC3659². L'adresse du serveur et les données de connexions sont fournies par la BnF. La BnF dispose d'une capacité réseau en entrant d'environ 1 Gbps à partir du réseau RENATER. Il appartient au prestataire de contracter le service réseau nécessaire au transfert des documents numériques.

¹ RFC959 < <http://www.ietf.org/rfc/rfc959.txt> >

² RFC3659 < <http://www.ietf.org/rfc/rfc3659.txt> >

4. DESCRIPTION GENERALE D'UN DOCUMENT NUMERIQUE

Un document numérique peut être constitué des éléments suivants :

- Le **fichier manifeste** du document numérique qui contient les métadonnées de production. Son contenu est décrit dans le « Référentiel d'enrichissement des métadonnées » pour les marchés fournissant des manifestes sous format refNum et dans le « Référentiel d'enrichissement des métadonnées – version METS » pour les autres ;
- Les **fichiers images** issus de la numérisation. Leur contenu est décrit dans les référentiels de prises de vue « Référentiel de numérisation des documents opaques » et « Référentiel de numérisation des documents transparents ». Le format des fichiers images est décrit dans le « Référentiel de format de fichier image » ;
- Eventuellement, un fichier image contenant la numérisation de la ou des **charte(s) colorimétrique(s)** retenue(s) pour le calibrage du numériseur qui a effectué la prise de vue. Leur contenu est décrit dans les référentiels de prises de vue (cf. point précédent « fichiers images ») ;
- Eventuellement, les **fichiers OCR** qui contiennent le résultat de la reconnaissance optique de caractères. Leur contenu est décrit dans le « Référentiel OCR » ;
- Eventuellement, le **fichier TdM** qui contient la transcription des pages contenant la ou les table(s) des matières. Son contenu est décrit dans le « Référentiel de traitement des tables » ;
- Eventuellement, le **fichier livre numérique (EPUB)** Son contenu est décrit dans le « Référentiel EPUB 3 ».
- Eventuellement, le fichier au **format pivot adaptable** décrit dans le « Référentiel DAISY ».

Selon le type de prestation, un paquet de livraison est constitué de tout ou partie de ces éléments.

5. ORGANISATION DES PAQUETS DE LIVRAISON

Les documents numériques sont livrés dans des fichiers d’empaquetage (paquet). Chaque paquet peut être accompagné d’un fichier de suivi contenant l’empreinte numérique du paquet permettant de vérifier son intégrité.

La livraison d’un ensemble de documents numériques est formalisée dans un fichier bordereau de livraison. Sa présence indique implicitement la fin de la livraison.

- Le **paquet** est un fichier au format zip. Son nom est constitué du code prestation (PPP) sur trois chiffres³ et de l’identifiant du document sur 7 ou 8 chiffres⁴ ;
- le **fichier de suivi** est un fichier texte contenant l’empreinte numérique (condensat) du paquet calculé selon l’algorithme MD5 (RFC1321⁵). Le nom du fichier de suivi est constitué du nom du paquet et de l’extension qui porte le nom de l’algorithme choisi (ici : .md5). ex : « PPP_1234567.zip.md5 »
- le **fichier bordereau de livraison** est un fichier au format XML tel que défini à l’adresse suivante :

<http://bibnum.bnf.fr/transmissionMessages/>

- les **fichiers manifestes** sont des fichiers XML respectant les schémas refNum⁶ ou METS⁷ ;
- le **fichier TdM** (table des matières) est un fichier XML au format tdmNum⁸ ;
- les **fichiers images** sont au format JPEG2000 ;
- la **charte colorimétrique** est un fichier image au format JPEG2000 sans perte ;
- les **fichiers OCR** sont des fichiers XML respectant le schéma alto_prod⁹ ;
- le **fichier format pivot adaptable** est conforme à la DTD DAISY XML (DTBook 2005-3). Ce format pivot pourra être amené à évoluer, vers la version courante de la DTD Daisy Digital Talking Book ou vers un autre schéma XML (ou DTD) apte à satisfaire les besoins de l’accessibilité¹⁰ ;
- le **fichier livre numérique** au format EPUB 3¹¹.



NOTES

La description détaillée des formats est indiquée dans les référentiels (cf. §4 Description générale d’un document numérique, plus haut).

³ Ce code est fourni par la BnF.

⁴ Il s’agit d’un numéro d’ordre complété à gauche par des zéros. La plage de numéros d’ordre utilisables est fournie par la BnF.

⁵ <http://www.ietf.org/rfc/rfc1321.txt>

⁶ <<http://bibnum.bnf.fr/refNum>>

⁷ <<http://www.loc.gov/standards/mets/>> et <<http://bibnum.bnf.fr/mets>>

⁸ <<http://bibnum.bnf.fr/tdmNum>>

⁹ Le schéma alto_prod <http://bibnum.bnf.fr/alto_prod> est une version BnF du schéma ALTO <<http://www.loc.gov/standards/alto>>.

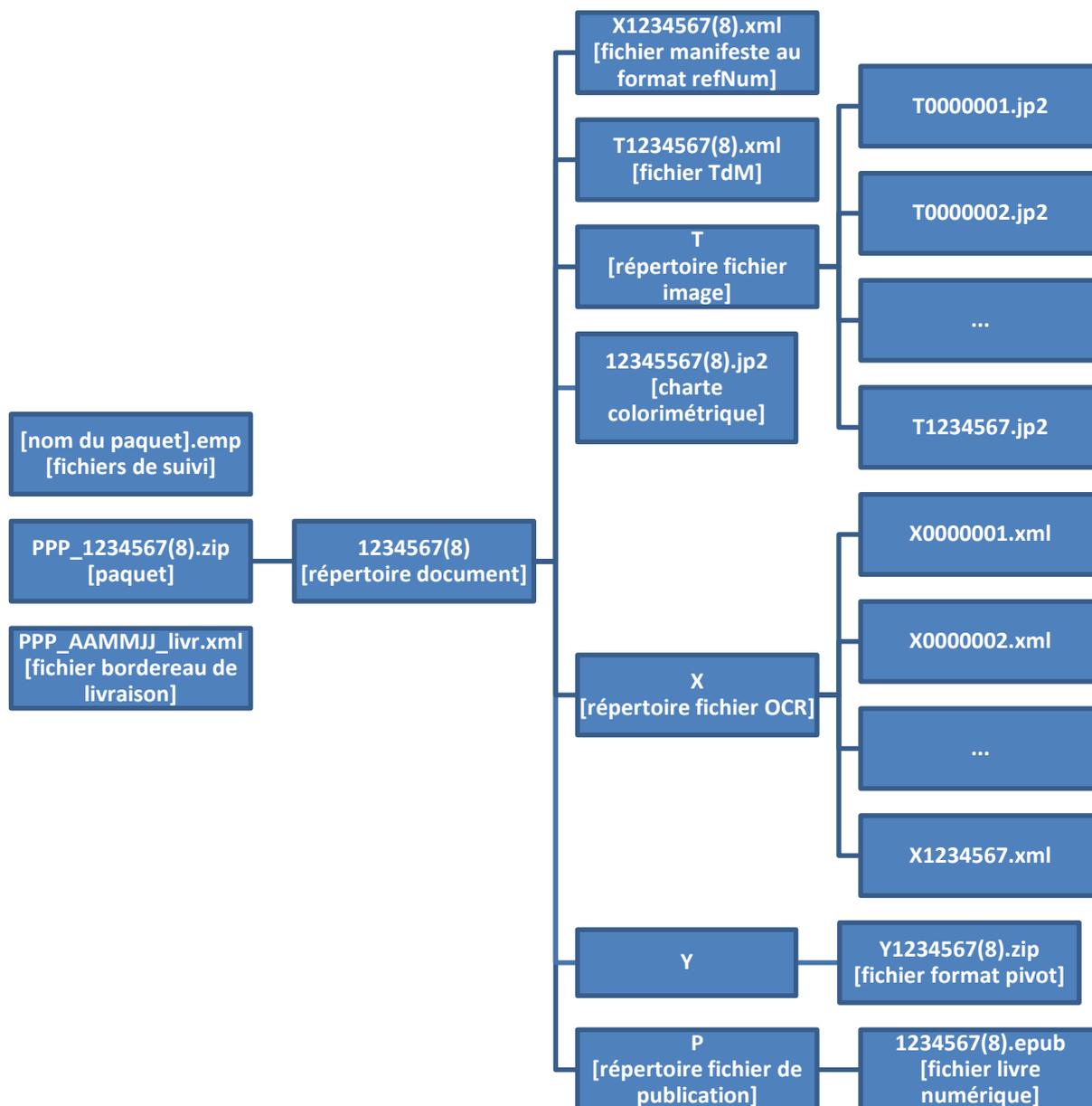
¹⁰ Par exemple ZedAI, http://www.daisy.org/zw/ZedAI_Introduction, ou TEI (<http://www.tei-c.org>).

¹¹ <<http://idpf.org/epub/30>>

Les longueurs des identifiants des documents numériques sont sur sept ou huit chiffres selon les marchés.

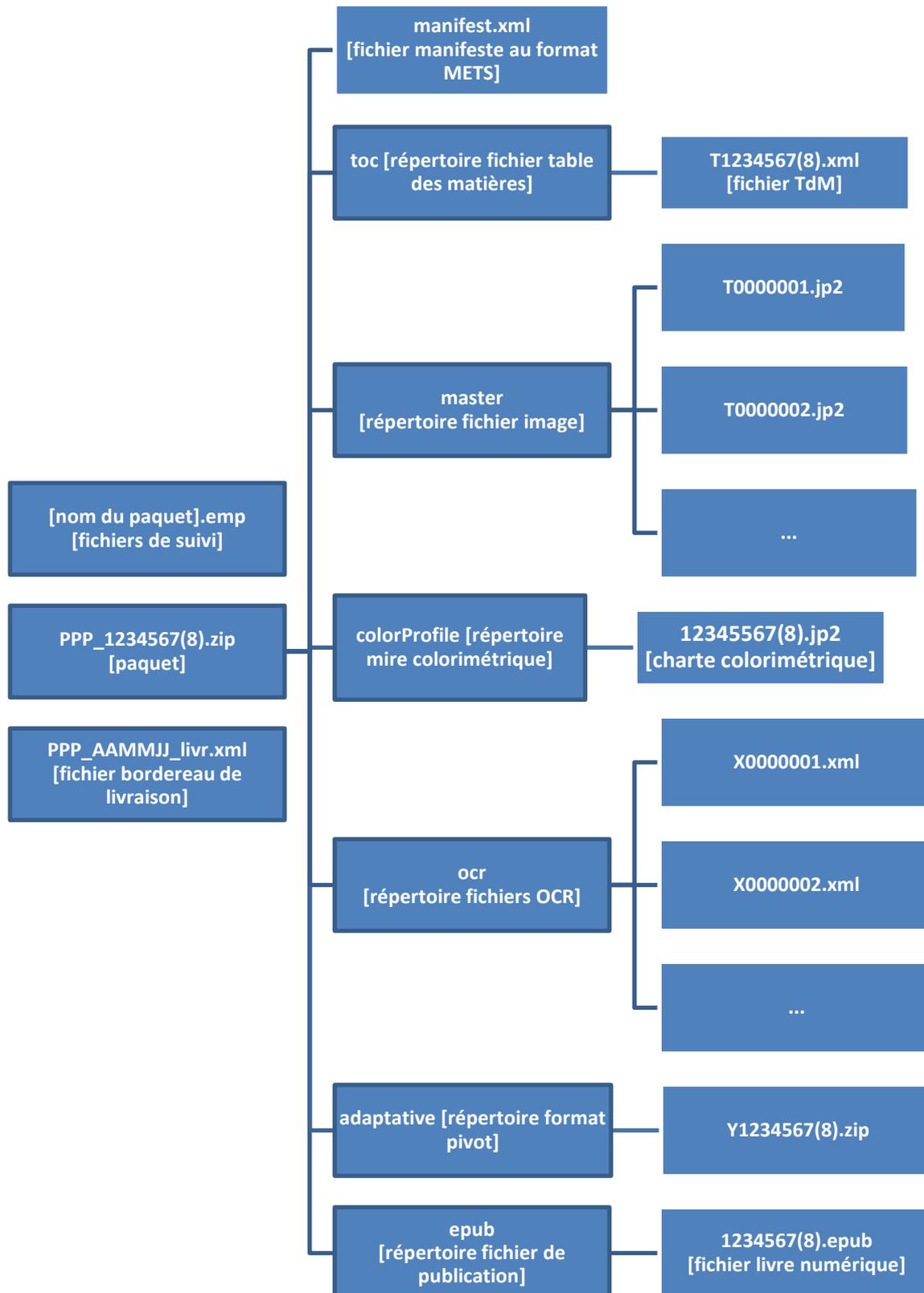
5.1 Organisation des livraisons - marchés RefNum

Les fichiers et les répertoires sont organisés de la manière suivante :



5.2 Organisation des livraisons – marchés METS

Les répertoires contenus dans le paquet zippé comprennent directement le fichier manifeste et des dossiers dont le nom correspond aux attributs USE des groupes de fichiers METS (éléments <fileGrp>s).



6. DESCRIPTION DES TYPES DE LIVRAISON

6.1 Livraison initiale

Ce type de livraison est utilisé lorsqu'un bordereau de traitement aller (BTA) a été fourni en accompagnement des objets physiques à numériser. Selon le type de prestation, deux structures sont attendues :

- les documents complets qui comprennent :
 - le fichier manifeste du document numérique,
 - les fichiers images,
 - les fichiers OCR,
 - le fichier TdM (si nécessaire),
 - le fichier EPUB (si demandé),
 - le fichier du format pivot (si demandé).
- les documents images seules qui comprennent :
 - le fichier manifeste du document numérique,
 - les fichiers images.

6.2 Livraison complémentaire

Ce type de livraison est utilisé lorsqu'un bordereau de traitement complémentaire (BTC) a été fourni en accompagnement des documents numériques à compléter. Selon le type de prestation plusieurs associations sont possibles. Le paquet de livraison contient a minima le fichier manifeste fourni par la BnF complété d'un ou plusieurs des éléments suivants :

- les fichiers OCR ;
- le fichier TdM (si nécessaire) ;
- le fichier EPUB (si demandé) ;
- le fichier du format pivot (si demandé).



Sur les marchés METS, des consignes spécifiques sont appliquées aux manifestes dans le cas de réfections courantes ou de traitements complémentaires. Cf. référentiel METS § 11.