

Référentiel de livraison de document numérique audiovisuel V1



<https://gallica.bnf.fr/ark:/12148/btv1b53161663s>

Table des matières

1	INTRODUCTION	3
2	DOCUMENTS APPLICABLES ET DE REFERENCE.....	4
3	GLOSSAIRE	5
4	DESCRIPTION DU SERVICE DE LIVRAISON	6
5	DESCRIPTION GENERALE D'UN DOCUMENT NUMERIQUE AUDIO	7
6	ORGANISATION DES PAQUETS DE LIVRAISON	8
6.1	Description du contenu de la livraison	8
6.2	Règles de nommage à respecter pour la livraison	9
6.2.1	Règle de nommage des éléments d'un paquet	9
6.2.2	Fichiers de numérisation et de métadonnées	10
6.2.3	Indication des faces des cassettes audio.....	11
7	DESCRIPTION DES TYPES DE LIVRAISON	13
7.1	Livraison initiale.....	13
7.2	Relivraison.....	13
8	ANNEXES	14
8.1	Exemple de manifeste.....	14
8.2	Exemple de fichier de métadonnées de gestion (MTD).....	14
8.3	Exemple de fichier de métadonnées techniques audio (MTA).....	15
8.4	Exemple de fichier de métadonnées techniques image (MTI)	16

1 INTRODUCTION

Ce référentiel s'inscrit dans le cadre des opérations de numérisation de documents audiovisuels réalisées pour le compte de la BnF.

Dans cette version, le référentiel ne traite que des documents audio. Une version ultérieure précisera les adaptations nécessaires pour traiter les cas des documents vidéo.

Le présent référentiel vise à regrouper les consignes à respecter pour la livraison des documents. Il sera enrichi au fur et à mesure des marchés entrepris pour la numérisation de documents audiovisuels.

De manière générale, les livraisons s'effectuent par dépôt sur un serveur FTP à l'aide du protocole sFTP.

2 DOCUMENTS APPLICABLES ET DE REFERENCE

En rapport avec les documents audiovisuels, les documents faisant référence aux critères de numérisation de la BnF sont disponibles soit en annexe au marché établi avec le prestataire, soit à cette adresse :

<https://www.bnf.fr/fr/les-referentiels-de-numerisation-de-la-bnf>

Intitulé	Description	Référence
<i>Annexe décrivant les métadonnées - version MT(x)</i>	<i>Document précisant pour chaque support audiovisuel, les informations à fournir et le format à respecter dans la saisie des informations.</i>	<i>Annexe au CCTP du marché</i>
<i>Annexe décrivant les conditions de numérisation des documents audio</i>	<i>Document indiquant le processus de numérisation pour chaque type de support constituant les documents audio et pris en compte dans les marchés passés avec la BnF.</i>	<i>Annexe au CCTP du marché</i>
<i>Référentiel de format de fichier audio</i>	<i>Document précisant les formats utilisés par la BnF dans le cadre de la numérisation de documents sonores en fonction de leur type et de leur particularité.</i>	https://www.bnf.fr/sites/default/files/2022-09/referentiel_format_fichier_audio_V1.pdf
<i>Format TIFF</i>	<i>Description du format TIFF v6 standard pour la prise de vues des supports, le cas échéant.</i>	<i>Standard format TIFF v6</i> https://web.archive.org/web/20210108172855/https://www.adobe.io/open/standards/TIFF.html
<i>Bordereaux</i>	<i>Fichiers permettant de déclarer un statut pour les documents (envoi, réception, livraison, contrôle...) et qui permettent de fournir des indications aux outils du circuit de la numérisation.</i>	http://bibnum.bnf.fr/transmissionMessages/ <i>Exemple complet de circuit d'un document numérisé :</i> http://bibnum.bnf.fr/transmissionMessages/exemples/circuit_5838768.html

3 GLOSSAIRE

- Le **paquet** est un fichier au format zip sans compression ou tout autre traitement complémentaire (par exemple le chiffrement). Son nom est constitué du numéro de la prestation (PPP) sur trois chiffres et de l'identifiant du document entre 6 et 9 chiffres ;
- Le **fichier d'empreinte** est un fichier texte contenant l'empreinte numérique (condensat) du paquet calculé selon l'algorithme MD5 (RFC 1321). Le nom du fichier d'empreinte est constitué du nom du paquet et de l'extension qui porte le nom de l'algorithme choisi (ici : .md5 en minuscule). Exemple : « PPP_1234567.zip.md5 » ;
- Le **fichier bordereau de livraison** est un fichier texte contenant une description de la livraison dans le langage XML tel que défini à l'adresse suivante : <http://bibnum.bnf.fr/transmissionMessages/> ;
- Le **fichier manifeste** est un fichier au format XML qui décrit le paquet (cf. section [Documents applicables et de référence](#)) ;
- Les **fichiers de métadonnées** sont des fichiers textes contenant une description des fichiers audio et images contenus dans le paquet utilisant un formalisme de type .INI (cf. section [Documents applicables et de référence](#)) ;
- Les **fichiers audio** sont au format FLAC, BWF ou DSD, selon les cas (cf. section [Documents applicables et de référence](#)) ;
- Les **fichiers images** sont au format TIFF V6 24 bits RGB non compressé (cf. section [Documents applicables et de référence](#)).

4 DESCRIPTION DU SERVICE DE LIVRAISON

Les documents numérisés doivent être livrés par réseau via un serveur FTP.
Le moyen de connexion utilisé est le protocole de transfert de fichiers sécurisé sFTP.

Ce protocole est conforme aux recommandations RFC959 et RFC4254.

L'adresse du serveur et les informations d'identification sont fournies par la BnF au prestataire en charge de la prestation de numérisation.

La BnF dispose d'une connexion à très haut débit à partir du réseau RENATER. Il appartient au prestataire de contracter le service réseau nécessaire au transfert des documents numériques pour lui permettre la transmission des livrables dans les meilleures conditions possibles.

5 DESCRIPTION GENERALE D'UN DOCUMENT NUMERIQUE AUDIO

Un document numérique audio peut être constitué des éléments suivants :

- Le fichier manifeste et les fichiers de métadonnées du document numérique qui contient les métadonnées de production. Leur contenu est précisé dans le [document de description des métadonnées MT\(s\) annexé au CCTP du marché concerné](#) ;
 - Les fichiers audio : leur contenu est décrit dans l'annexe au CCTP du marché concerné. Le format des fichiers audio est décrit dans le « [Référentiel de format de fichier audio](#) » ;
 - Et éventuellement des fichiers images : le format des fichiers images est conforme au standard TIFF v6 24 bits RGB non compressé.
- Ces éléments sont une partie constituante du paquet de livraison.

6 ORGANISATION DES PAQUETS DE LIVRAISON

Les documents numériques audio (cf. section [Description générale d'un document numérique audio](#)) sont livrés sous la forme de fichiers d'empaquetage (paquet – cf. section [Glossaire](#)).

En plus du contenu correspondant au document, chaque paquet contient un fichier manifeste permettant entre autres à la BnF de faire le lien entre le document de support d'origine et le document sous sa forme numérisée.

Le paquet sera accompagné de son empreinte numérique pour permettre de vérifier son intégrité.

Dans un second temps, la livraison d'un ensemble de documents numériques audio est formalisée dans un fichier bordereau de livraison (cf. section [Documents applicables et de référence](#)). Le dépôt de ce fichier indique la fin de la livraison des paquets venant d'être chargés sur le serveur FTP.

6.1 Description du contenu de la livraison

A chaque document correspond une cote (référence unique BnF) qui peut être constituée d'un ou plusieurs volumes (ou supports ou cassettes) identifiés par la mention Vn_m. Les informations de volumes sont données par la BnF sous forme d'un Bordereau de Traitement Aller (BTA). Il s'agit d'un fichier texte contenant un descriptif du lot dans un langage XML.

Les livraisons pour les marchés Son doivent avoir le contenu décrit ci-après, un paquet (paquet – cf. section [Glossaire](#)).

Ce paquet sert à regrouper tous les éléments découlant de la numérisation d'un support audio, à savoir :

- Un manifeste (fichier manifeste – cf. section [Glossaire](#)) ;
- Le répertoire du volume du document contenant :
 - Les fichiers de métadonnées audio par numérisation répondant au [document annexe](#) précisant le schéma de l'enrichissement des métadonnées MT(x). Le nombre de ces fichiers est fonction du support numérisé ;
 - Un ou plusieurs fichiers de numérisation audio en fonction du support, du nombre de faces... ;
 - Un ou plusieurs fichiers de métadonnées image du support en fonction du type de support numérisé, répondant au [document annexe](#) précisant le schéma de l'enrichissement des métadonnées MT(x) ;
 - Un ou plusieurs fichiers de numérisation image en fonction du support, du nombre de faces...

6.2 Règles de nommage à respecter pour la livraison

6.2.1 Règle de nommage des éléments d'un paquet

Ce sous-chapitre s'appuie sur les 3 sous-chapitres précédents, à savoir que chaque type de répertoires ou de fichiers a son appellation normée, notamment pour permettre aux outils du circuit de numérisation (CNUM) de prendre en compte les documents attendus. Par conséquent, les outils du CNUM rejeteront les paquets ne répondant pas aux règles citées ci-après.

Les règles de nommage sont définies de la manière suivante (les termes entre crochets doivent être remplacés et adaptés en fonction du contexte de la prestation commandée) :

- Le bordereau de livraison **\$PPP_[123456789]_[AAMMJJ]_livr.xml** ;
- Le paquet **\$PPP_[123456789].zip** ;
- Le fichier **\$PPP_[123456789].zip.md5**.

A l'intérieur du paquet **\$PPP_[123456789].zip** :

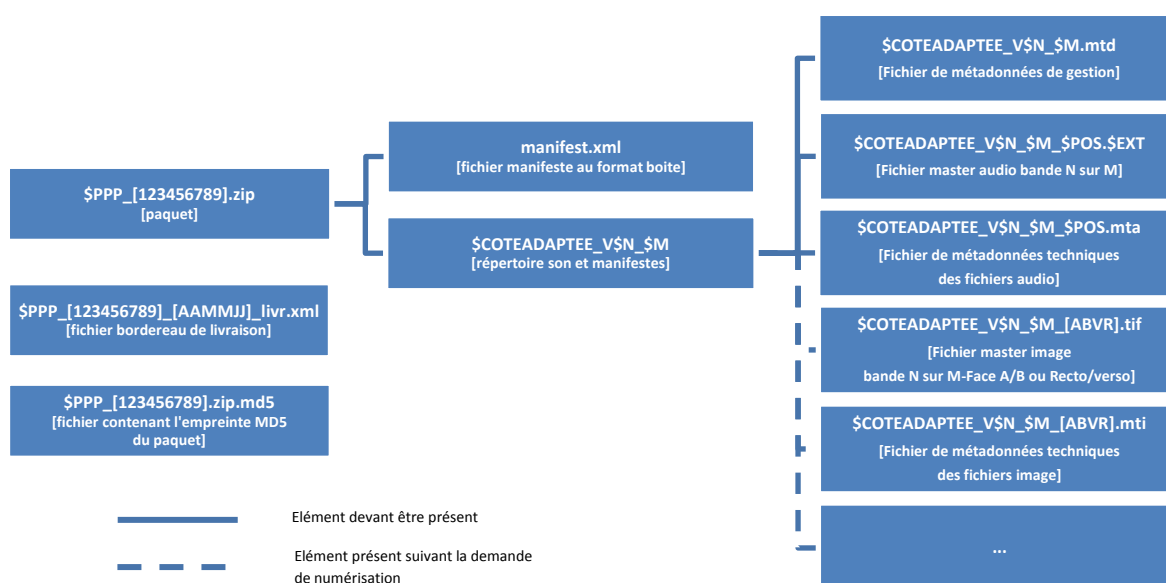
- Le manifeste manifest.xml (ce fichier doit impérativement avoir cette dénomination) ;
- Le répertoire du volume **\$COTEADAPTEE_V\$N_\$M** ;
- Le manifeste de gestion **\$COTEADAPTEE_V\$N_\$M.mtd** ;
- Les manifestes techniques audio **\$COTEADAPTEE_V\$N_\$M_\$POS.mta** ;
- Les fichiers audio **\$COTEADAPTEE_V\$N_\$M_\$POS.\$EXT** ;
- Les manifestes techniques pour les fichiers images **\$COTEADAPTEE_V\$N_\$M_[ABRV].mti** ;
- Les fichiers images **\$COTEADAPTEE_V\$N_\$M_[ABRV].tif**.

→ Glossaire d'appellation :

- **\$PPP** : correspond au numéro de la prestation attribué au lot de document à la suite de l'attribution du marché ;
- **123456789** : permet d'assigner un identifiant unique au document numérique. Celui-ci peut comporter jusqu'à 9 chiffres. Des tranches d'identifiants seront fournies au prestataire pour qu'il puisse en doter chaque document numérique ;
- **AAMMJJ** : indique la date du bordereau (AA pour le nombre des dizaines de l'année, MM correspondant au numéro du mois et JJ pour le jour d'émission du bordereau) ;
- **\$COTEADAPTEE** : est une adaptation de la cote du document physique qui peut porter une notion de « variante ». Pour prendre un exemple : pour la cote SM-1, on notera SM_000001. Chaque espace est supprimé. Le trait d'union est remplacé par un tiret bas (représenté par le caractère « _ »). La numérotation de la cote doit comporter 6 chiffres. Exemple 2 : pour la numérisation de la cote DONAUD1714-2 dans sa variante en DSD, on notera DONAUD1714_000002_(MASTER-DSD). De façon générale, la BnF portera à la connaissance du prestataire les cotes à numériser et les correspondances associées (méthode d'adaptation de la cote et variante) ;
- **V\$N_\$M** : détermine la position du document dans chaque volume numérisé. Le caractère « V » sera suivi d'un premier chiffre indiquant la position (**\$N**), tandis que le second chiffre (**\$M**) fournira le nombre total de volumes lié à la cote traitée. Un document n'ayant qu'un seul support physique se verra systématiquement marqué « V1_1 » ;

- **\$POS** : mentionne une valeur révélant la présence pour un document de plusieurs possibilités de lecture (présence de différentes faces, valeurs : A ou B, de plusieurs bobines, valeurs : A, B, C, D... ; de plusieurs plages, valeurs : 001, 002, 003, ...)
- **[ABRV]** : précise pour la prise de vue s'il s'agit du recto (**R**) ou du verso (**V**) du support photographié pour une jaquette et il s'agit des faces **A** ou **B** pour les cassettes audio ;
- **\$EXT** : correspond à l'extension du fichier en fonction des critères de numérisation (support, format...). Cette variable peut avoir la valeur .flac, .wav et .dsd (cf. [Référentiel de format de fichier audio](#)).

Ci-dessous un exemple du formalisme commun appliqué selon les règles citées précédemment.



6.2.2 Fichiers de numérisation et de métadonnées

Communément, chaque répertoire de volume permettra d'obtenir les fichiers suivants conformément au référentiel indiqué à la section [Glossaire](#) :

- Des fichiers audio au format de fichier correspondant au support traité par face, le cas échéant ;
- Des fichiers de métadonnées techniques audio et de gestion ;
- Des fichiers images correspondant à la numérisation du support, le cas échéant ;
- Des fichiers de métadonnées techniques image.

Pour la lecture de la règle de nommage dans l'exemple ci-après, il faut se reporter à la section 6.2.1.

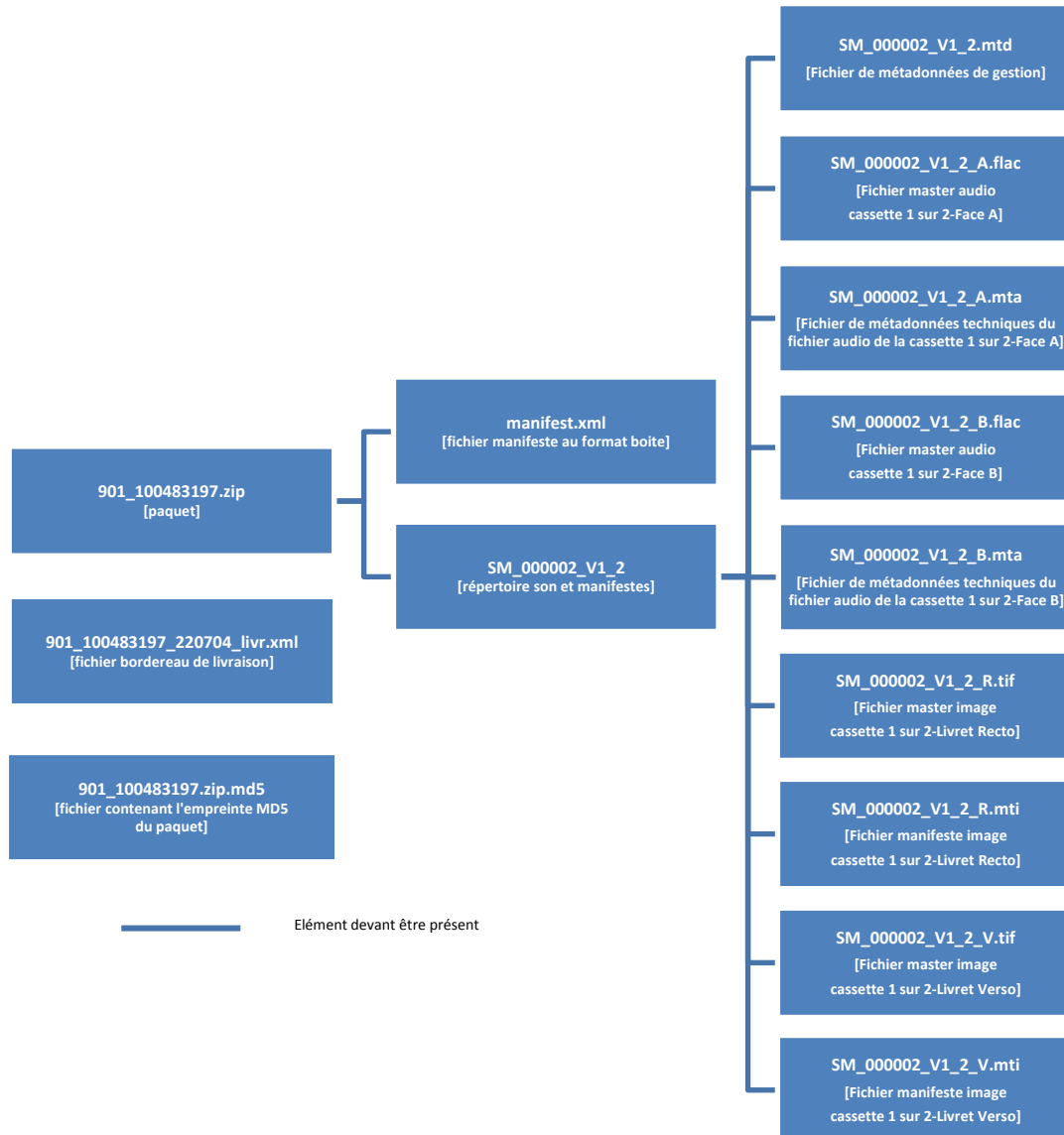


EXEMPLE

Pour la cote SM-2 contenant 2 volumes, on notera pour le volume 1 : SM_000002_V1_2.

Le document de cette cote est une cassette audio ayant 2 faces.

L'exemple de numérisation s'effectue sur la prestation 901 et l'identifiant attribué au document numérique est 100483197.



7 DESCRIPTION DES TYPES DE LIVRAISON

Ce chapitre fait appel à la nomenclature des bordereaux dont certains exemples sont fournis au lien indiqué à la section [Documents applicables et de référence](#).

7.1 Livraison initiale

Ce type de livraison est utilisé lorsqu'un BTA a été fourni en accompagnement des objets physiques à numériser.

Comme évoqué au chapitre organisation des paquets de livraison et selon le type et les particularités du document, un contenu est attendu et il est accompagné d'un Bordereau de Livraison (BL).

7.2 Relivraison

Ce type de livraison est utilisé dans deux cas de figure :

- lorsqu'un bordereau de contrôle a été émis par les outils de contrôle automatique à la suite d'une détection d'une malfaçon dans les fichiers attendus par le circuit de numérisation ;
- lorsqu'un bordereau de contrôle d'admission totale (BCAT) a été fourni pour commander une reprise de document à la suite d'un contrôle qualité, comme précisé dans le CCTP du marché.

C'est uniquement par la réception de l'un de ces deux bordereaux, que le prestataire peut de nouveau livrer les documents pour lesquels une demande de réfection a été indiquée. La livraison s'effectuera de nouveau avec un Bordereau de Livraison (BL).



NOTES

Des consignes spécifiques sont appliquées aux manifestes dans le cas de réfections courantes.

8 ANNEXES

8.1 Exemple de manifeste

Fichier XML en UTF-8 sans BOM

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<boite nom="100_800001">
<!-- identifiant de production du paquet -->
  <identification id="800001">
    <titre>Audio test FLAC</titre>
  </identification>
  <structure>
    <volume no="1" sur="1"></volume>
    <repertoire>SDC12_045039_V1_1</repertoire>
  </structure>
  <liens>
    <notice_BnF>ark:/12148/cb382384025</notice_BnF>
<!-- identifiant du document = identifiant de production du paquet volume 1 -->
    <id_document>800001</id_document>
    <cote_originale>SDC 12-45039</cote_originale>
  </liens>
  <contenu>
    <item no="1" type="fichier"
taille="1395">SDC12_045039_V1_1\SDC12_045039_V1_1.mta</item>
    <item no="2" type="fichier"
taille="377">SDC12_045039_V1_1\SDC12_045039_V1_1.mtd</item>
    <item no="3" type="fichier"
taille="225687960">SDC12_045039_V1_1\SDC12_045039_V1_1.flac </item>
  </contenu>
</boite>
```

8.2 Exemple de fichier de métadonnées de gestion (MTD)

Fichier texte en UTF-8 sans BOM suivant un formalisme de type .INI

```
[DTD]
version= AUDIO 1.00
[DOCUMENT]
titre= UNISSONS - Flûte japonaise 2e transpo.
volumaison= 1 sur 1
cote_ex_original= DONAUD1714_002447_V1_1
cote_ex_numerise=
[SUPPORT]
support_ex_original =
support_ex_numerise=
support_ex_securite=
[NUMERISATION]
```

etat =
[Validation]
etat =

8.3 Exemple de fichier de métadonnées techniques audio (MTA)

Fichier texte en UTF-8 sans BOM suivant un formalisme de type .INI

[METADONNEES-AUDIO]

VERSION 1.02

[DOCUMENT]

Titre=RFI Brosseau, Fabienne

Statut du titre=Certain

Provenance / fonds=

Cote BnF du support copié=ASPSON 012789

Référence d'inventaire présente sur le support copié=

Marque de l'original=

Référence commerciale=

Volume=1/1

Face=A

Type de cassette=Compact Cassette

Marque et type de support enregistrable=RFI

Type de support=Fer

Durée cassette vierge=30

Type de signal=Stéréo

Nombre de pistes=2

Réduction de bruit=Aucune

Collure=non

Etat de conservation=Bon

Observation conservation=

[CREATION OBJET NUMERIQUE]

Nom de fichier=ASPSON_012789_V1_1_A.wav

Durée totale (mm:ss:ff)=00:40:35

Date de transfert (jj/mm/aaa)=10/12/2018

Prestataire de transfert=STUDIO X

Nettoyage préalable=non

Lecteur=TASCAM 122 MK3 s/n 0900606

Vitesse (cm/s)=4,75

Courbe d'égalisation=IEC Type I

Réduction de bruit lors du transfert=non

Type de transfert=Stéréo

Convertisseur=PRISM SOUND AD2 s/n 2612352187

Résolution (bits)=24

Echantillonnage (kHz)=96

Edition=Wavelab 7.2.1 (build 600) - 32 bit

Gain numérique=0

Format de fichier=FLAC

Checksum MD5=f25a56542731730657712008b5cb616a

Fades entrée-sortie=oui

Restauré=non

Prestataire restauration=

Date de restauration (jj/mm/aaaa)=

Observation restauration=

Etat vérification=

[OBSERVATION SIGNAL AUDIO]
Souffle=oui
Usure=non
Pleurage=non
Ronflette=non
Saturation=non
Copie ou dub=
Qualité auditive (1 à 10)=8
Observation signal=Début tronqué conforme à l'original - Distorsion sur certaines sibilantes

[DECOUPAGE]
Nombre total de plages=1
Début et durée plage 1 (mm:ss:ff/mm:ss:ff)=00:00:00/00:40:35

8.4 Exemple de fichier de métadonnées techniques image (MTI)

Fichier texte en UTF-8 sans BOM suivant un formalisme de type .INI

```
[METADONNEES-IMAGE]
VERSION 1.00

[DOC]
Titre=RFI Brosseau, Fabienne
Cote BnF du support copié=ASPSON 012789
Volume=1/1
Type de support copié=Compact Cassette
Etat de conservation=
Observation conservation=

[CREATION OBJET NUMERIQUE]
Nom image=ASPSON_012789_V1_1_R.tif
Taille image (octets)=7229444
Date de prise de vue (jj/mm/aaaa)=07/12/2018
Titulaire prise de vue=STUDIO X
Nettoyage préalable=non
Dispositif de prise de vue=Scanner EPSON V500 PHOTO s/n K65W071154
Type de prise de vue=Scan
Logiciel de prise de vue=SilferFast 8.0
Référence de calibration=MIRESCAN_20181207_B01_STUDIOXK7.tif
Profil ICC=RVB: Adobe RGB (1998)
Résolution (dpi)=400
Profondeur de codage (bits)=24
Format de fichier=TIFF V6 non compressé
Réglage Gamma=Gris: Gray Gamma 2.2
Etalonnage=KodakIT8.7
Compression=non compressé
Edition=non
Retailage=non
Watermarking=non
Etat vérification=

[OBSERVATION IMAGE]
Qualité visuelle (1 à 10)=
Observation image=
```