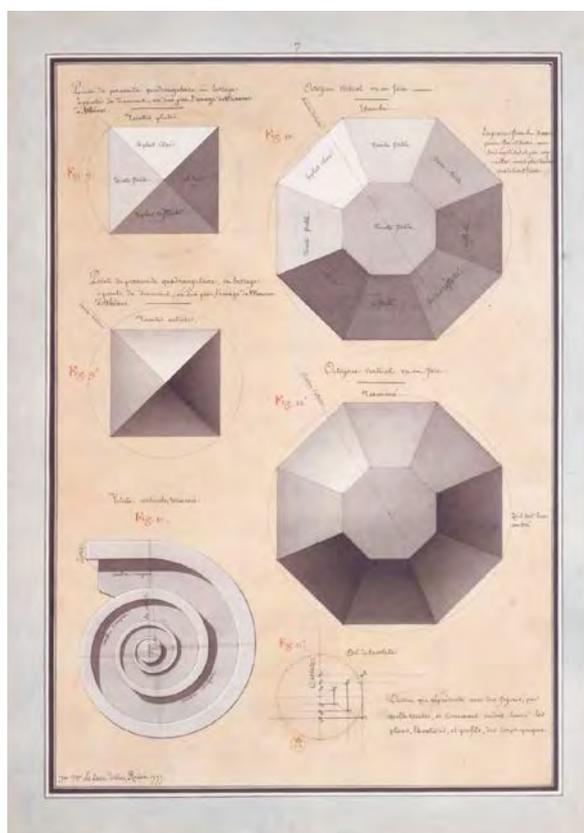


Référentiel de format de fichier image v2



Source gallica.BnF / Bibliothèque nationale de France

Bibliothèque nationale de France
direction des Services et des réseaux
département de la Conservation
service Numérisation

Date : le vendredi 24 avril 2015
Version : 1
Référence BnF : BnF-ADM-2014-071949-02

TABLE DES MATIERES

1.	INTRODUCTION	3
1.1	CONTEXTE ET OBJECTIF	3
1.2	GESTION DE VERSION	3
2.	DESCRIPTION DU FORMAT	4
2.1	FORMAT DE FICHIER	4
2.2	COMPRESSION	4
2.3	PROFIL COLORIMETRIQUE	5
2.4	RESOLUTION	5
2.5	NIVEAU DE QUALITE	5
2.6	ORDRE DE PROGRESSION	5
2.7	ROBUSTESSE DU CODE JPEG2000	5
3.	OUTILS DE REFERENCE ET QUALIFICATION DU PROCEDE DE PRODUCTION	6
4.	METADONNEES	7
4.1	METADONNEES DESCRIPTIVES	7
4.2	METADONNEES DE CREATION D'IMAGE	8
4.3	METADONNEES D'HISTORIQUE DE TRAITEMENT	8

1. INTRODUCTION

1.1 Contexte et objectif

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 création ou génération de fichiers images issus de la numérisation de documents physiques ou d'objets.

Depuis les premiers programmes de numérisation en 1993, la BnF a fait le choix d'utiliser le format TIFF V6. D'abord utilisé en multipages (plusieurs images dans un même fichier image), ce format fut, à partir de 1996, utilisé dans une version monopage (une image par fichier image).

Depuis 2009, la BnF étudie l'opportunité de passer au format JPEG2000. La complexité du format, le faible nombre d'outils fiables et le nombre restreint d'utilisateurs étaient jusqu'à présent considérés comme autant de risques et de freins à l'adoption de ce format.

A ce jour, ces freins demeurent, notamment quant à la faiblesse de l'offre logicielle capable de traiter correctement ce type de format d'image. Néanmoins, le nombre d'utilisateurs a largement augmenté ; nombre de bibliothèques ont adopté ce format : Library of Congress, British Library, Koninklijke Bibliotheek, etc.

De plus, la parution de la *ISO/IEC 15444-1:2004/Amd 6:2013*, permet de clarifier d'une part, l'utilisation des profils ICC avec la possibilité d'utiliser des profils de type 'DISPLAY' et, d'autre part, l'utilisation des champs : « capture resolution » et « default display resolution ».

Pour la BnF, les avantages offerts par ce format sont indéniables :

- Capacité de compression sans perte visuelle importante ;
- Nombreuses fonctionnalités pour la diffusion : capacité d'extraire rapidement un niveau de résolution sans nécessité de générer des images de résolutions intermédiaires ;
- Forte expressivité des métadonnées.

1.2 Gestion de version

Version	Réf. BnF	Date de validation	Résumé des modifications et ajouts
V0	BnF-ADM-2013-079325-01	07.11.2013	Première version. Pas de prise en compte des documents transparents et des documents de la presse.
V1	BnF-ADM-2014-071949-01	02.12.14	Précisions sur les métadonnées et sur les paramètres de la commande <code>kdu_compress</code> .
V2	BnF-ADM-2014-071949-02	en cours	Ajout compression des documents transparents et méthode de test du ratio de compression.

2. DESCRIPTION DU FORMAT

Cette description concerne également les images « charte ».

2.1 Format de fichier

Le format de fichier retenu est JPEG2000 issu des normes ISO 15444-1 et 15444-2 et leurs correctifs et compléments.

2.2 Compression

Le type de compression privilégié est une compression irréversible (lossy). Il s'agit de la transformation en ondelette 9-7 flottant, **ICT** (« *irreversible component transform* »).

Le niveau de compression dépend du type de collection. Il est exprimé par un ratio cible par rapport à la taille (poids informatique de l'image en octets) de l'image originale dans le tableau suivant :

Type Collection	Ratio de compression cible
Documents « spécialisé », estampes, photographie, plaque de verre, etc. numérisés en couleur (24 bits par pixel).	1:4
Documents « exceptionnel », imprimés ou manuscrit comportant des enluminures, des illustrations ou des dépliants particuliers numérisés en couleur (24 bits par pixel).	1:4
Documents « imprimé » numérisés en couleur (24 bits par pixel).	1:6
Documents « transparent » numérisés en niveaux de gris (8 bits par pixel).	1:16
Documents de la « presse » grand format numérisés en niveaux de gris (8 bits par pixel).	nc ¹

Le ratio compression cible obtenu avec l'algorithme de compression irréversible (transformation en ondelette 9-7 flottant, ICT) est dépendant de la densité informationnelle de l'image à compresser. Ainsi, il n'est pas possible de prévoir le ratio après compression.

Une marge de tolérance est fixée à 5%.

Afin de se prémunir de toute perte d'information excessive et non contrôlée. Les images avec un ratio de compression supérieur à la tolérance doivent être compressées selon la transformation en ondelette 5-3 entier, **RCT** (*Reversible Component Transform*), dite compression sans perte.

¹ La BnF n'a pas déterminé le ratio de compression pour ce type de collection.

2.3 Profil colorimétrique

Le profil colorimétrique tel que décrit dans les référentiels images doit être présent dans le fichier.

Il est inscrit dans la « Colour Specification » Box de l'en-tête JP2 en utilisant une description « Restricted ICC » de classe « Display Device Profile ».

2.4 Résolution

La résolution de capture telle que décrite dans les référentiels images² doit être présente dans le fichier.

Elle est inscrite dans une « Capture resolution box » de l'en-tête JP2 et exprimée en « pixel per meter ».

Le nombre de niveau de résolution requis est 10.

2.5 Niveau de qualité

Le nombre de niveau de qualité requis est 10.

2.6 Ordre de progression

L'ordre de progression requis est RPCL.

2.7 Robustesse du code JPEG2000

Afin d'obtenir un code JPEG2000 robuste, il est demandé d'intégrer les marqueurs suivants :

- SOP (Start Of Packet)
- EPH (End of packet header)
- PLT (packet length, tile-part header)

² cf. Référentiel de numérisation des documents opaques et Référentiel de numérisation des documents transparents

3. OUTILS DE REFERENCE ET QUALIFICATION DU PROCEDURE DE PRODUCTION

L'outil de référence pour vérifier la conformité des fichiers JPEG2000 est jpylyzer dans sa version 1.10.1 ou supérieure :

<http://www.openplanetsfoundation.org/software/jpylyzer>

L'outil de référence pour la production de fichier JPEG2000 est kakadu dans sa version 7.2.1 ou supérieure.

<http://www.kakadusoftware.com/>

La BnF doit valider le paramétrage utilisé pour générer les images au format JPEG2000. Les images brutes issues de la numérisation ne doivent pas avoir été compressées avant la transformation au format JPEG2000.

Si le matériel de capture (scanner ou camera) le permet, les images peuvent être directement produites au format JPEG2000. Dans ce cas, la validité des images produites par ce procédé sera qualifiée par la BnF à partir d'images de test et charte de couleur standard en comparaison avec la chaîne de production suivante :

Image brute de scan au format TIFF non compressé → transformation au format JPEG2000 avec « kdu_compress » (kakadu).

Exemple de ligne de commande avec kdu_compress en demandant un ratio de compression de 1:4³ pour un fichier TIFF couleur 24 bits non compressé :

```
kdu_compress \  
-i <fichierEntree.tif> -o <fichierSortie.jp2> \  
-rate 6.0 Clayers=10 \  
Creversible=no Clevels=10 \  
Stiles={1024,1024} Cblk={64,64} \  
Corder=RPCL \  
ORGtparts=R \  
Cprecincts={256,256},{256,256},{128,128} \  
Cuse_sop=yes Cuse_eph=yes ORGgen_plt=yes \  
-jp2_box <fichierUUIDbox> \  
-precise
```

où fichierUUIDbox est un fichier binaire commençant par la ligne « uuid » et contenant l'information nécessaire pour produire les métadonnées définies ci-dessous

³ Pour atteindre un ratio de compression de 1:4, il est nécessaire d'imposer un débit de 6.0 bits/échantillon dans le cas d'un fichier codé sur 24 bits.

4. METADONNEES

Le format de métadonnées conteneur utilisé est XMP tel que décrit dans la norme ISO 16684-1:2012 et son extension disponible sur :

<http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMPSpecificationPart2.pdf>.

Conformément à la spécification XMP Part 3⁴, cette description XMP est inscrite dans le fichier JP2 sous forme d'une box UUID d'identité :

0xBE7ACFCB97A942E89C71999491E3AFAC.

Les métadonnées doivent respecter ces normes. En outre, les règles de nommage et gestion des préfixes des namespaces doivent être appliquées. On veillera à encapsuler la description XMP par une directive `<?xpacket?>` précisant, dans l'attribut `begin`, l'encodage UTF-8 de cette description pour éviter notamment une mauvaise interprétation des caractères accentués.

De manière générale, les métadonnées générées par le matériel de capture⁵ et, le cas échéant, l'outil de génération du fichier doivent être conservées. De plus, il convient d'enregistrer des informations descriptives de base, ainsi que des informations d'historique sur la création du fichier image décrites dans les chapitres suivants. Ces métadonnées sont obligatoires. Le format XMP fonctionne comme un format conteneur réutilisant des vocabulaires pré-existants. Aussi, les vocabulaires à utiliser pour chaque type de métadonnées sont précisés.

4.1 Métadonnées descriptives

Des informations minimales sur le document numérique doivent être exprimées dans chaque fichier JPEG2000. L'objectif est d'avoir une information la plus stable possible. Ces informations utiliseront le vocabulaire Dublin Core simple (préfixe "dc:") et le vocabulaire Dublin Core qualifié (préfixe "dcterms:"). Il convient de se référer au « référentiel d'enrichissement des métadonnées » qui donne plus de précision sur ces métadonnées.

Les éléments à renseigner sont :

- L'identifiant de production du document numérique, exprimé sous la forme « NUM » ou « IFN », underscore (« _ »), puis un numéro d'ordre à l'intérieur d'une plage de valeurs clairement identifiée. Cette information est exprimée dans un champ **dcterms:isPartOf**.
- Le propriétaire du document numérique. Il s'agit *a minima* de « Bibliothèque nationale de France ». Cette information est exprimée dans un champ **dcterms:provenance**. Elle peut être complétée par un second champ **dcterms:provenance** portant la mention du partenaire dont le document numérique est issu.
- L'identifiant ARK de la notice bibliographique du document. Cette information est exprimée dans un champ **dc:relation**.

⁴

<http://www.adobe.com/content/dam/Adobe/en/devnet/xmp/pdfs/XMPSpecificationPart3.pdf>

⁵ Caméra ou scanner

- La cote de l'original reproduit (il s'agit de la cote de la Partie exemplaire correspondante). Cette information est exprimée dans un champ **dc:source**.

4.2 Métadonnées de création d'image

Ces métadonnées utilisent les vocabulaires XMP (qui correspondent au noyau de métadonnées XMP tel qu'exprimé dans la partie 1 de la norme, préfixe : "xmp:") et les métadonnées spécialisés TIFF (documentées dans la partie 2 de la norme, préfixe : "tiff:"), EXIF (documentées dans la partie 2 de la norme, préfixe : "exif:") et AUX (documentées dans la partie 2 de la norme, préfixe : "aux:")

Les éléments à renseigner sont :

- La date de création du fichier numérique⁶. Cette date doit être exprimée au format ISO-8601 dans la propriété **xmp:CreateDate**.
- La date de dernière modification du fichier numérique⁷. Il s'agit de la date de la dernière manipulation ayant modifié le fichier avant envoi à la BnF. Cette date doit être exprimée au format ISO-8601 dans la propriété **xmp:ModifyDate**.
- L'outil matériel (scanner ou appareil photo) utilisé pour la numérisation ou prise de vue, correspondant au nom du modèle. Cette information est exprimée dans la propriété **tiff:Model**.
- Le constructeur du matériel décrit en **tiff:Model**, exprimé dans une propriété **tiff:Make**.
- Le numéro de série du matériel décrit en **tiff:Model**, exprimé dans une propriété **aux:SerialNumber**
- Le constructeur, le nom et la version du logiciel de numérisation ou de création du fichier, exprimé dans une propriété **xmp:CreatorTool**.
- Le nom du prestataire de numérisation effectuant l'opération, exprimé dans la propriété **tiff:Artist**.

Les dates sont exprimées avec une précision ne dépassant pas la seconde.

Ces métadonnées sont les seules obligatoires. D'autres champs, notamment les champs EXIF, peuvent être renseignés dans l'image sans que cela soit imposé.

4.3 Métadonnées d'historique de traitement

Ces métadonnées concernant le post-traitement des images, sont à exprimer dans les métadonnées extérieures aux fichiers JPEG2000, au format refNum et METS/PREMIS, telles que exprimées dans le « référentiel d'enrichissement des métadonnées ».

⁶ Il s'agit donc de la date de création initiale du fichier au format JPEG2000. Elle ne doit pas évoluer même en cas de modification.

⁷ Il s'agit donc de la date de la dernière modification du fichier au format JPEG2000.



Dans cet exemple :

- l'identifiant de production du document est NUM_6127578
- la notice bibliographique correspondante est ark:/12148/cb328608203
- l'exemplaire original ayant servi à la numérisation est FOL-Z-219
- le fichier image a été initialisé le 28 juin 2010 à 9h36 et modifié pour la dernière fois le 29 juin 2010 à 17h35
- le scanner utilisé est un Scanstation Rollfilm de la société Wicks and Wilson, dont le numéro de série est 952840341659
- le logiciel de traitement d'image utilisé est SpiFactory 3.2.12 de la société SPIGRAPH
- le prestataire de numérisation est PRESTATAIRE_NUM.

L'attribut begin contient le caractère Unicode U+FEFF encodé en UTF-8 (soit la séquence 0xEF 0xB 0xBF)

```
<?xpacket begin="\x{FEFF}" id="W5M0MpCehiHzreSzNTczkc9d"?>
<x:xmpmeta xmlns:x="adobe:meta/">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description
      xmlns:xmp="http://ns.adobe.com/xap/1.0/"
      xmlns:tiff="http://ns.adobe.com/tiff/1.0/"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:dcterms="http://purl.org/dc/terms/"
      xmlns:exif="http://ns.adobe.com/exif/1.0/"
      xmlns:aux="http://ns.adobe.com/exif/1.0/aux/"
      rdf:about=""
      tiff:Model="Scanstation Rollfilm"
      tiff:Make="Wicks and Wilson"
      tiff:Artist="PRESTATAIRE_NUM"
      aux:SerialNumber="952840341659"
      xmp:CreatorTool="SPIGRAPH SpiFactory 3.2.12"
      xmp:CreateDate="2010-06-28T09:36:00Z"
      xmp:ModifyDate="2010-06-29T17:35:00Z">
      <dcterms:isPartOf>
        <rdf:Bag>
          <rdf:li>NUM_6127578</rdf:li>
        </rdf:Bag>
      </dcterms:isPartOf>
      <dcterms:provenance>
        <rdf:Bag>
          <rdf:li>Bibliothèque nationale de France</rdf:li>
        </rdf:Bag>
      </dcterms:provenance>
      <dc:relation>
        <rdf:Bag>
          <rdf:li>ark:/12148/cb328608203</rdf:li>
        </rdf:Bag>
      </dc:relation>
      <dc:source>
        <rdf:Bag>
          <rdf:li>FOL-Z-219</rdf:li>
        </rdf:Bag>
      </dc:source>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
<?xpacket end="r"?>
```