institut national d'histoire de l'art

INHA

ANNEXE 1 - PLAN DE NOMMAGE DES FICHIERS NUMERIQUES

BIBLIOTHEQUE NUMERIQUE DE L'INHA

OBJECTIF : garantir l'accessibilité du document, éviter les problèmes techniques, faciliter la recherche et assurer une bonne gestion des documents

Pas de normes officielles (cf. http://www.enssib.fr/content/regles-de-nommage-des-fichiers-et-repertoires) mais des règles de bonnes pratiques :

* Utiliser pour chaque fichier **un nom unique**. Cela permet **d'éviter une confusion entre les fichiers et leurs différentes versions** (par exemple palier le problème des fichiers dégradés pour diffusion portant le même nom mais rangés dans des répertoires différents).
* Le nom d'un fichier doit être **succinct** : il ne doit pas dépasser 31 caractères sans compter l'extension, mais **précis** : type de document, cote...
* Le nom d'un fichier ne doit pas contenir : d'espace, de ponctuation (sauf pour l'extension), de caractères accentués ou spéciaux (ùé+'@à°[ ] :</* »& !$, etc)**, utiliser uniquement les lettres a…z et les chiffres 0…9** - Seul le signe _ (underscore) est autorisé et recommandé pour distinguer des entités au sein du nom du fichier (mais attention en cas d'utilisation sur le web l'underscore peut être confondu avec le soulignement propre au lien hypertexte)

L'ensemble des fichiers (nommés livrables) à fournir, leur nommage et l'organisation de leur répertoire sont précisés lors des réunions de projet propres à chaque fonds.

NOMMAGE DES FICHIERS SOURCES

La majorité des fichiers de l'INHA compte entre 30 et 34 digits (idéalement 31).

Exemples:

Livres anciens : **B751025206C000008VA1536**_071.tif ;
B751025206C008PIECE2595_001.jpg
Photographies : **B751025206PH0000PF23_01**_008.tif
Catalogues de vente numérisés d'après l'original :
B751025206CVPXXXXX_AAAAMMJJ_001.tif
Catalogues de vente numérisés d'après l'original en interne (cote à la place du n° d'ordre) : **B751025206CVP1940-105_AAAAMMJJ**_001.tif
Catalogues de vente 20e siècle numérisés d'après le microfilm :**B751025206CVXXXXX_AAAAMMJJ**_001.jpg
Cartons verts : **B751025206CVERTG_DRUET_1903-1909_001**.tif
Autographes numérisés d'après microfilm : **B751025206AUMF01C001**_000009.jpg
Autographes numérisés d'après l'original mais disposant d'une cote microfilm :
B751025206AUORMF03C007_002037-002038.tif
Autographes numérisés d'après l'original ne disposant pas d'une version microfilm :
B751025206AUORAUT211_10_001.tif
Autographes numérisés dans les collections BCMN : **B751015202AUOR0630_02_01**_001.tif
Manuscrit : **B751025206MS0000000MS81**_482.tif
Périodiques : **B751025206P13U4_1903-1909**_001.tif

• **B751025206** est invariable et représente le numéro identifiant de la Bibliothèque de l'INHA dans le SUDOC (nº RCR), **B751015202** pour la BCMN.
• **XXXX** est invariable et indique le fonds numérisé, il peut être formé de 1 à 4 caractères selon les cas : **AR** pour archives
AUMF pour autographes (numérisé d'après microfilm)
AUOR pour autographes (numérisé d'après l'original)
C pour classiques de l'histoire de l'art
CV pour catalogue de vente (numérisé d'après microfilm)
CVA pour cartons verts Artistes
CVG pour cartons verts Galeries
CVP pour catalogue de vente (numérisé d'après l'original)
D pour dessin
E pour estampe
MS pour manuscrit
PH pour photographie

• **XXXXXXXXXXX** est la cote normalisée des documents originaux ou le numéro d'ordre attribué au catalogue de vente pour servir d'identifiant unique. Cette cote normalisée est éventuellement précédée de zéros pour atteindre les 31 caractères). Pour les cartons verts, une cote normalisée est attribuée par pochette et le nommage des fichiers des pièces contenues dans une même pochette se fera dans la continuité d'une pièce à l'autre.
• _ est un séparateur.
[• Pour les catalogues de vente : **AAAAMMJJ** est la date de la vente sui vi du séparateur _]
• 001 est le numéro séquentiel de chaque vue numérisée à l'intérieur du document.
• .jpg ou .tif est l'extension du fichier produit.

NOMMAGE DES FICHIERS POUR LA MISE EN LIGNE

Modèle de nommage fichier PDF

Pour la mise en ligne les fichiers de diffusion seront mis au format

- JPEG pour les images de documents en feuille (photographie, estampe, dessin, feuillet de lettres),
- PDF pour les recueils (livres, périodiques, recueils d'estampes, catalogues de vente, documents recto verso) ; ces fichiers PDF contenant les images au format JPEG et (la plupart du temps) une table des matières arborescente intégrée pourront être des PDF multicouches avec le mode texte résultant de l'OCR (optical character recognition) encapsulé.

Le but est d'appliquer le moins de compression possible (200 dpi pour les PDF) mais les PDF sont enregistrés en version 1.4 (Acrobat 5.x) et leur poids n'excède pas 2 Go par fichier en raison des impératifs de la plateforme de diffusion.

Le nommage des fichiers de diffusion comporte la rubrique de classement dans la bibliothèque numérique (code fourni pour chaque fonds par le pouvoir adjudicateur au titulaire), la bibliothèque de provenance et la cote du document.

Exemple : **1053**_**doucet**_**4PHOT15_02**.pdf

- **1053** : code de rubrique de classement dans la bibliothèque numérique
- _ : séparateur
- **doucet** : provenance de l'original
- _ : séparateur
- **4PHOT15_02** : cote normalisée
- .pdf : extension

Les tables des matières sont élaborées par l'équipe du SID. Le nommage de leurs fichiers (xlsx et txt) est identique à celui du fichier de diffusion sans sa rubrique de classement, par exemple : doucet_4PHOT15_02.xlsx et doucet_4PHOT15_02.txt

Modèle de nommage fichier XML Alto

Même règle que pour les fichiers TIFF avec extension en .xml. Les numéros séquentiels de page doivent correspondre entre les fichiers TIFF et les fichiers XML. Ex. : B751025206C000000004I97_011.tif correspond à B751025206C000000004I97_011.xml = OCR de la vue n° 11

Modèle de nommage fichier MD5

Même règle que pour les fichiers TIFF mais en s'arrêtant après la cote normalisée (puisqu'il s'agit d'un fichier unique) et avec l'extension en .md5. Il faut ajouter le nombre de 0 nécessaire pour arriver à 31 caractères.