CAHIER DES CHARGES POUR LA NUMERISATION DES MANUSCRITS

DESCRIPTION DU MATERIEL

IRHT-CNRS
Institut de recherche et d'histoire des textes
Service Images - Pôle numérique
https://www.irht.cnrs.fr

Sommaire :

1. Objectif et rappel des contraintes
2. Conditions matérielles de la numérisation
3. État de conservation avant la numérisation et manipulation des ouvrages
4. Éclairage
5. Installation et description du matériel
6. Typologie des documents et modes de reproduction
7. Étalonnage
8. Gestion de la couleur
9. Réglage de la température de couleur
10. Exposition et contrôle qualité des images
11. Format de fichier, échantillonnage et résolution
12. Préparation des documents avant la numérisation
13. Organisation des fichiers numériques
14. Métadonnées
15. Archivage pérenne
16. Traitement et intégration des manuscrits dans la BVMM et dans la base Medium
17. Livraison des fichiers

1. Objectif et rappel des contraintes

Recommandations techniques pour la numérisation de manuscrits et d'imprimés anciens.
L'étude suivante a pour objet de définir la liste du matériel nécessaire, les protocoles de numérisation et les étapes de traitement des images.

Rappel des contraintes : matériel transportable, pas de lumière continue compte tenu de la fragilité des documents, image de qualité professionnelle.

2. Conditions matérielles de la numérisation

Utilisation d'un matériel adapté à la nature physique des documents : manuscrits reliés ou feuillets détachés. Le support sur lequel sont placés les documents doit permettre d'en préserver l'intégrité. La numérisation du document sera réalisée livre ouvert à 180 ° lorsque la reliure et l'ouverture des cahiers le permettent. Dans le cas contraire, le manuscrit ne sera ouvert qu'à 100° de façon à préserver la reliure et le montage des cahiers, utilisation d'un numériseur en berceau BC 100.

En aucun cas, le document ne sera plaqué sous une vitre sans le recours à des plateaux compensateurs.

Le presse-livre doit être spécifiquement conçu pour la reproduction des documents précieux.

Seul le responsable du fonds patrimonial pourra déterminer si le document peut être traité avec une ouverture à 180° ou à 100°. L'IRHT mettra en œuvre l'installation d'un matériel adéquat : numérisation en double page (résolution de sortie : 6496 x 4872 pixels et ou 6208 x 8280 pixels) ou dispositif permettant une prise de vue page à page avec une ouverture à 100° (numériseur en berceau).

Le photographe devra manipuler les documents originaux avec le plus grand soin et de la meilleure façon possible. Il utilisera uniquement le crayon pour prendre des notes (les taches d'encre moderne étant indélébiles) et il le posera lorsqu'il faut tourner le feuillet.

3. État de conservation avant la numérisation et manipulations

Recommandations :
Rédaction d'un bordereau pour signaler et décrire l'état de la reliure et des cahiers avant la prise de vue.
Consulter le manuscrit avec les mains propres.
Manipuler le livre sur un pupitre ou à plat sur un bureau.
Soutenir l'ouverture sans la forcer.
Manipuler avec soin les anciennes reliures : ouvrir et fermer doucement.
Ne jamais poser un livre sur un autre livre ouvert.
Ne jamais toucher la surface enluminée ou écrite, même avec une règle.
Afin de suivre un texte ligne à ligne, utiliser une feuille de papier A4 pliée en deux.

Pour collationner les cahiers d'un manuscrit dont la reliure est serrée ou en mauvais état, ne jamais approcher le visage trop près (le parchemin ou le papier réagit considérablement aux changements d'humidité).

4. Éclairage de prise de vue

Les sources lumineuses doivent être spécialement étudiées pour ne pas nuire à la préservation des documents (décoration et supports d'une grande fragilité). L'éclairage doit-être le plus faible possible et sans effet de chaleur. Il doit être d'une puissance constante et totalement paramétrable.

Les éclairages à lumière continue sont à éviter, excepté les lumières dites froides (LED). La température de couleur doit être de type lumière du jour équilibrée à **5500 K°.** Elle doit être constante et doit pouvoir être régulièrement contrôlée par l'opérateur depuis le logiciel de capture. L'éclairage doit également être le plus court possible. Le flash de studio est préconisé car l'éclair est très bref et sa puissance totalement modulable. C'est ce système d'éclairage (flashs équipés de boite à lumière de studio pour adoucir la lumière) qui est utilisé par l'IRHT depuis 1992.

Les lumières dites continues peuvent être utilisées notamment les caissons LED car ils ne dégagent aucune chaleur. Les autres sources peuvent être utilisées sous réserve qu'elles n'éclairent le document que durant le passage du CCD. Les technologies Tri CCD linéaire qui utilisent un néon ou une source de puissance supérieure à **220 lux** et qui passent à 3 cm au-dessus du document sont à proscrire, les temps de numérisation étant généralement trop longs. Avant de procéder à la numérisation, le bon fonctionnement du matériel installé sera vérifié et des tests devront être effectués à l'aide d'objets non sensibles autres que les manuscrits.

5. Installation et description du matériel

Station classique (prise de vue bi feuillet – verso + recto)

L'installation du matériel nécessite quelques contraintes, un local d'environ 10 m^2, une hauteur de plafond de 2,3 m, la possibilité d'occulter les fenêtres pour éviter le mélange de lumières et les reflets, l'accès à des prises de courant. Le système d'éclairage est disposé à 45 ° au-dessus des documents et à distance d'un mètre.

Croquis de l'installation

Exemple d'une installation à la médiathèque de Rennes Métropole

La station de numérisation comporte :

- 1 ordinateur
- 1 Hasselblad P 30 plus Mo
- 1 optique zoom HC 55/110mm
- 1 optique 120mm macro
- 1 pied photo colonne
- 2 flashs de studio + boîtes à lumière + pieds
- 1 presse-livre conçu pour la reproduction des documents rares et précieux et/ou

un module permettant l'ouverture de l'ouvrage à 100 °.

Caractéristiques technique de la numérisation :

- Prise de vue au format RAW
- Type de capture RVB Matricielle dos Phase One P 30
- Taille des fichiers : 6496 pixels x 4872 pixels ou supérieur 6208 x 8280 pixels
- Echantillonnage 16 bits par couleurs primaires
- Température de couleur : 5600 K° flash électronique de studio
- Insertion d'un Profil ICC sRGB et ou Adobe RGB 1998)
- ISO 100
- Capture One version DB 12 ou 20 64 bits

Numériseur en Berceau BC 100 (prise de vue page à page)

La structure de numérisation en berceau BC-100 est prévue pour numériser de manière rapide, fiable et qualitative des ouvrages reliés ou des documents simples.
Son berceau en V, dont l'angle d'ouverture est de 100° est ajustable. Il permet de numériser des documents dont les pages peuvent mesurer jusqu'à 432mm x 610mm. Le berceau peut être réglé en fonction de la largeur de la reliure, d'une largeur maximum de 15 cm (6 pouces). La manipulation reste en permanence sous le contrôle de l'opérateur. Cela permet, si besoin, d'ajuster le système afin d'assurer l'intégrité du matériel et de garantir une qualité d'image constante. Le changement de page est effectué manuellement, par l'opérateur. Le berceau est monté sur amortisseurs pneumatiques. Cette option permet de régler précisément la pression appliquée sur l'ouvrage afin de préserver les ouvrages anciens et fragiles. De même, le dos du berceau est recouvert de mousse haute densité afin de protéger les différents éléments de la reliure.
L'éclairage est assuré par des caissons lumineux équipés de LED équilibrés pour la lumière du jour (5500°K). Ces derniers assurent un éclairage uniforme sur toute la

surface de numérisation. La structure est fournie avec une tente en tissu pour éviter les lumières parasites et un compresseur adapté de qualité professionnel.

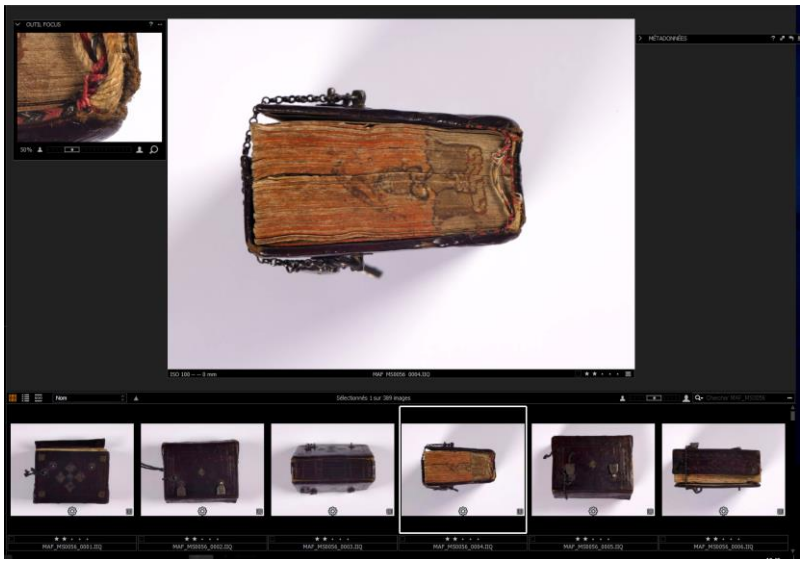
Installation du BC 100 à la bibliothèque d'Arras

L'installation nécessite une surface au sol de 8 m^2 minimum et une hauteur de 2,3 m.

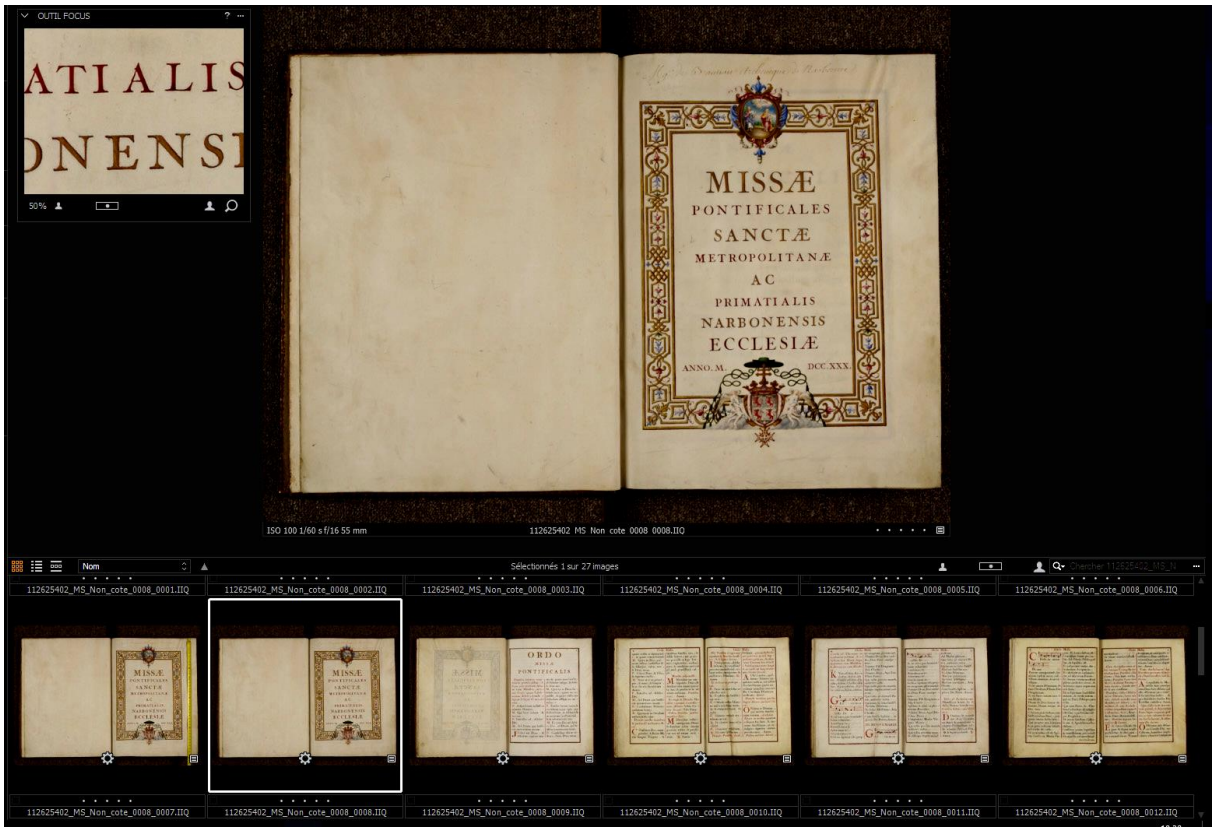
La capture est assurée par un ensemble de système PHASE ONE XF + IQ150, d'une résolution de 50 millions de pixels chacun. Deux jeux d'optiques (2x 80mm et 2x 120mm) permettent de couvrir tous les formats entre le A2 et le A5. Ces équipements professionnels disposent de la fonction live view afin de faciliter les opérations de cadrage et de mise au point.

6. Typologie des documents

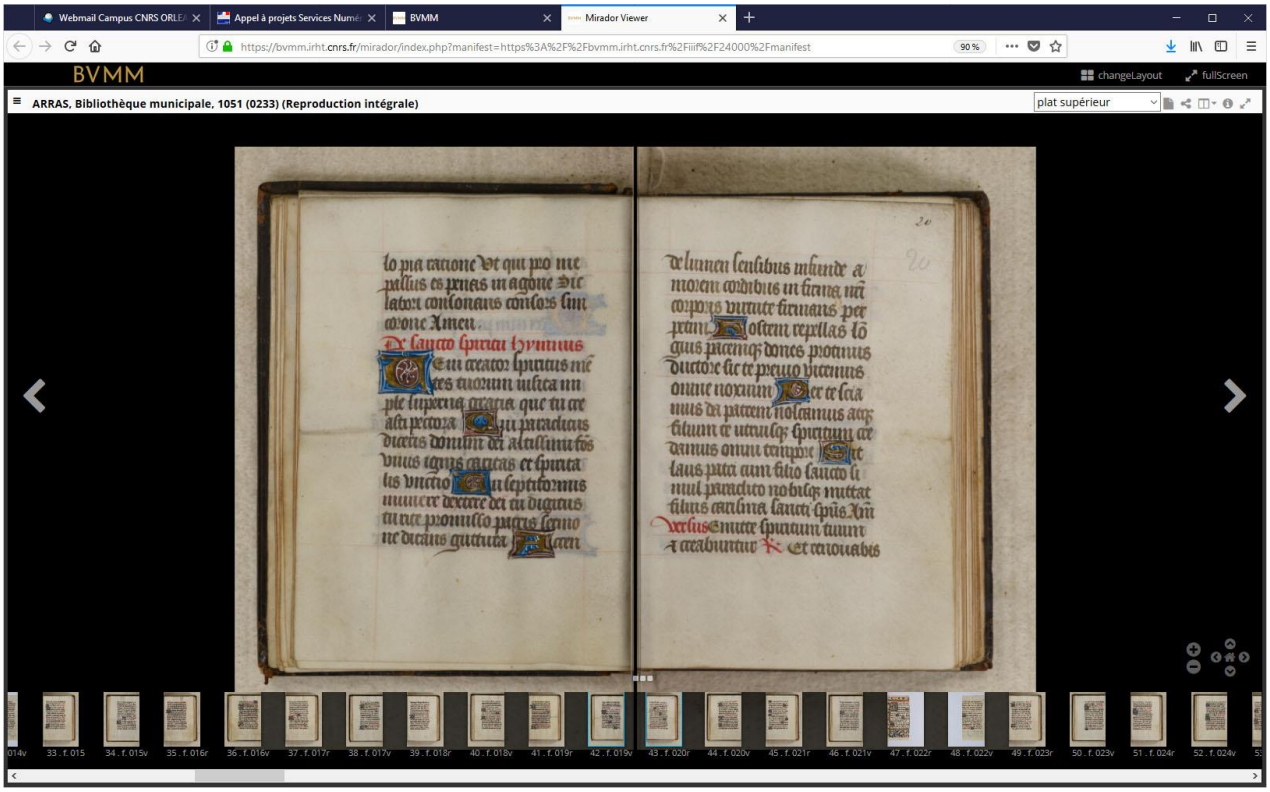
Toutes les parties du document sont à numériser :
La reliure, les plats, le contreplat sup. et inf., le dos, les tranches, les gardes, les feuillets vierges et les feuillets comportant le texte manuscrit.



La prise de vue s'entend livre ouvert (bi-feuillet : verso et recto du document). Lorsque la reliure est fragile et ne permet pas une ouverture à 180°, les prises de vue sont de type mono-page (1 prise de vue par page : recto et verso).



Prise de vue bi-feuillet



Prise de vue page à page – Viewer Mirador et serveur IIIF

7. Étalonnage

Avant le lancement de la production, il est nécessaire de réaliser un **étalonnage de la station** ainsi qu'un test effectué sur quelques folios du manuscrit.

L'étalonnage du moniteur

S'assurer de la bonne restitution des couleurs implique une vigilance sur l'ensemble de la chaîne de traitement : la gestion de la couleur est particulièrement complexe, elle nécessite un réglage en amont de l'ensemble des périphériques.
L'étalonnage du moniteur est une étape importante pour garantir une **visualisation correcte** des prises de vue numériques. Il permet d'enregistrer au cœur même du système d'exploitation un **profil ICC** correspondant à votre écran. Cet étalonnage doit être refait régulièrement.
Le moniteur de la station est étalonné à partir d'une sonde de type **Xrite Sonde i1 Display Pro** (spectrophotomètre), en vue d'obtenir une visualisation neutre des couleurs.
Réglage physique des luminophore RVB de l'écran (**profil Srvb***) avec un Gamma de 2,2.
Choisir un gamma de 2,2 dans le profil-écran vous garantit une bonne reproduction des dégradés.
Plus d'informations sont disponibles sur le site suivant :

http://www.profil-couleur.com/pp/306-gamma-ecran.php
Sans ce réglage indispensable réalisé en début de campagne, l'opérateur risque de travailler en aveugle et de faire, s'il ne prend pas soin de vérifier l'histogramme du fichier, des images **surexposées** ou **sous-exposées**.

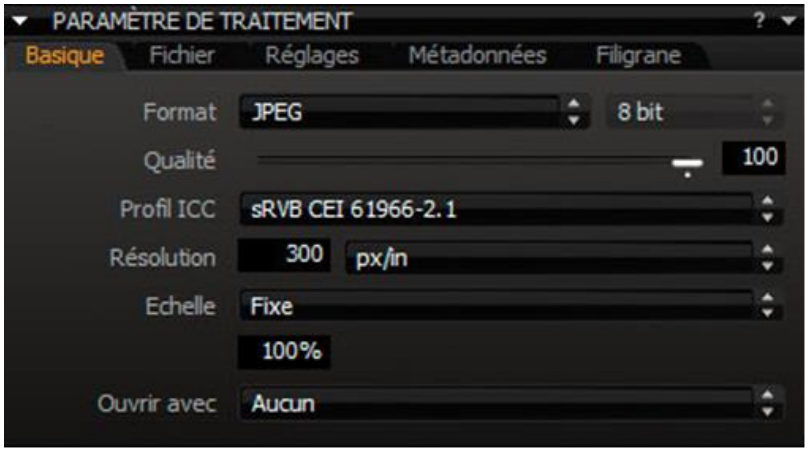
La température de couleur de d'écran est réglée à **9300 Kelvin** elle peut être établie à partir de la **sonde Xrite Sonde i1 Display Pro**

Interface de l'étalonnage du moniteur

8. Gestion de la couleur

L'échantillonnage de l'appareil permet une acquisition, au minimum, en RVB en 12 bits par couleur primaire.
Le système d'acquisition comporte un profil d'entrée ICC ou **ICM** (International Color Consortium) de façon à caractériser l'espace colorimétrique de l'appareil de prise de vue et un profil de visualisation standard

Le profil sRVB ou Adobe RGB (1998) est recommandé (standard rouge vert bleu) au moment de l'export Tif, Jpg ou Jpg 2000



L’IRHT utilise les logiciels d’acquisition **Phocus** et ou **Capture One Pro,** ces applications professionnelles permettent une gestion efficace et normalisée de la couleur et intègrent une gestion cohérente des profils ICC* en entrée et en sortie.

Les profils ICC* sont des fichiers permettant de convertir les couleurs d’un espace colorimétrique d’un périphérique à l’espace supporté par un autre périphérique, par exemple convertir les couleurs en Rouge, Vert, Bleu (RVB) en Cyan, Magenta, Jaune, Noir (CMJN) pour l’impression ou pour une consultation sur le web en utilisant le profil Srgb.

***ICC** : International Color Consortium

9. Réglage de la température de couleur

Le photographe établit une valeur numérique étalon des couleurs, valable pour toute la suite de la numérisation.

Le contrôle de cet étalonnage s’effectue pour chacun des documents à numériser.

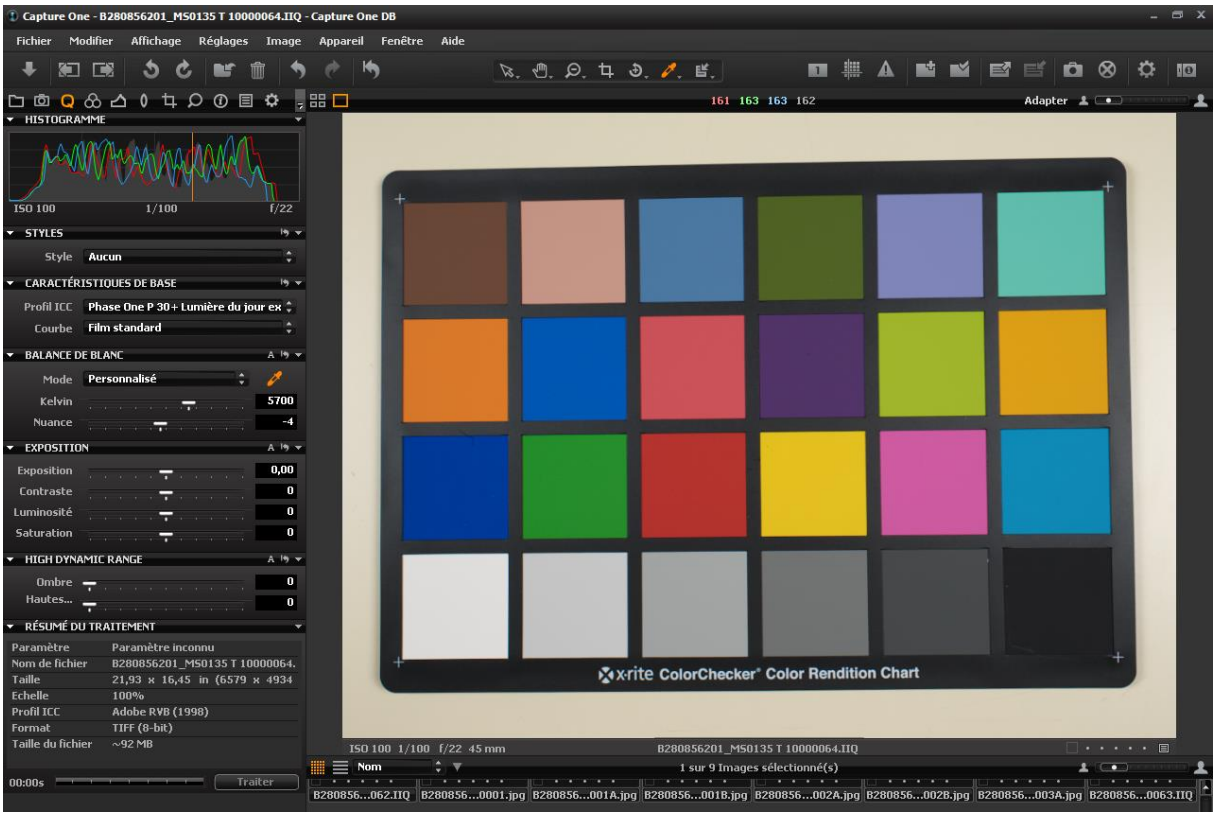
La valeur numérique étalon des couleurs établie reste valable pour toute la suite de la production. Cet étalonnage est réalisé à partir d’une charte couleur de type Gretag Macbeth.

L'équilibrage de la température de couleur est réglée en fonction de la source de l’éclairage utilisé : flash électronique équilibrée pour la lumière du jour à 5500 K° ou tungstène 3200 K°, il peut être également ajusté soit à partir du logiciel d’acquisition en utilisant un étalon blanc ou un gris neutre.

Les systèmes d'acquisition disposent d'un outil de mesure appelé pipette pour contrôler les références RVB des couleurs de la charte ou d'une zone neutre.

En réalisant un point gris sur cette plage, l'opérateur obtient un codage RVB des niveaux identiques et neutraliser ainsi une dérive de la température de couleur.

R : 173
V : 173
B : 173



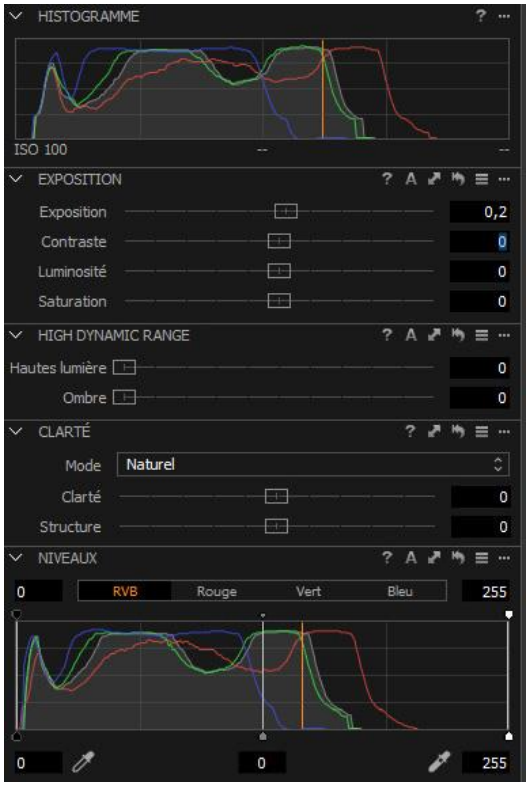
Contrôle de la température de couleur avec l'application Capture One Pro

10. Exposition

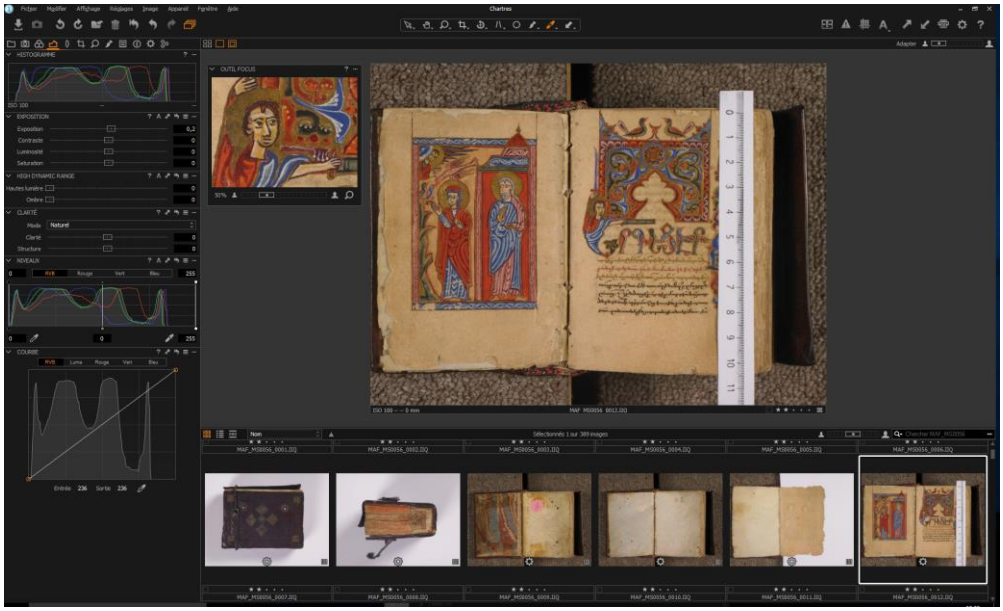
Le contrôle de l'exposition et des niveaux de luminosité est réalisé à partir de l'histogramme. Le photographe utilise trois paramètres pour obtenir une exposition correcte :

- Couple : diaphragme/vitesse
- Sensibilité du capteur en privilégiant une faible sensibilité entre 50 et 400 Iso.
- La puissance de l'éclairage et ou du flash électronique

Il est procédé au contrôle de **l'histogramme** des images, durant toutes les étapes de l'acquisition, en veillant à l'homogénéité des niveaux de luminosité sur les **trois couches** RVB et ce pour chaque prise de vue, tout au long du processus.

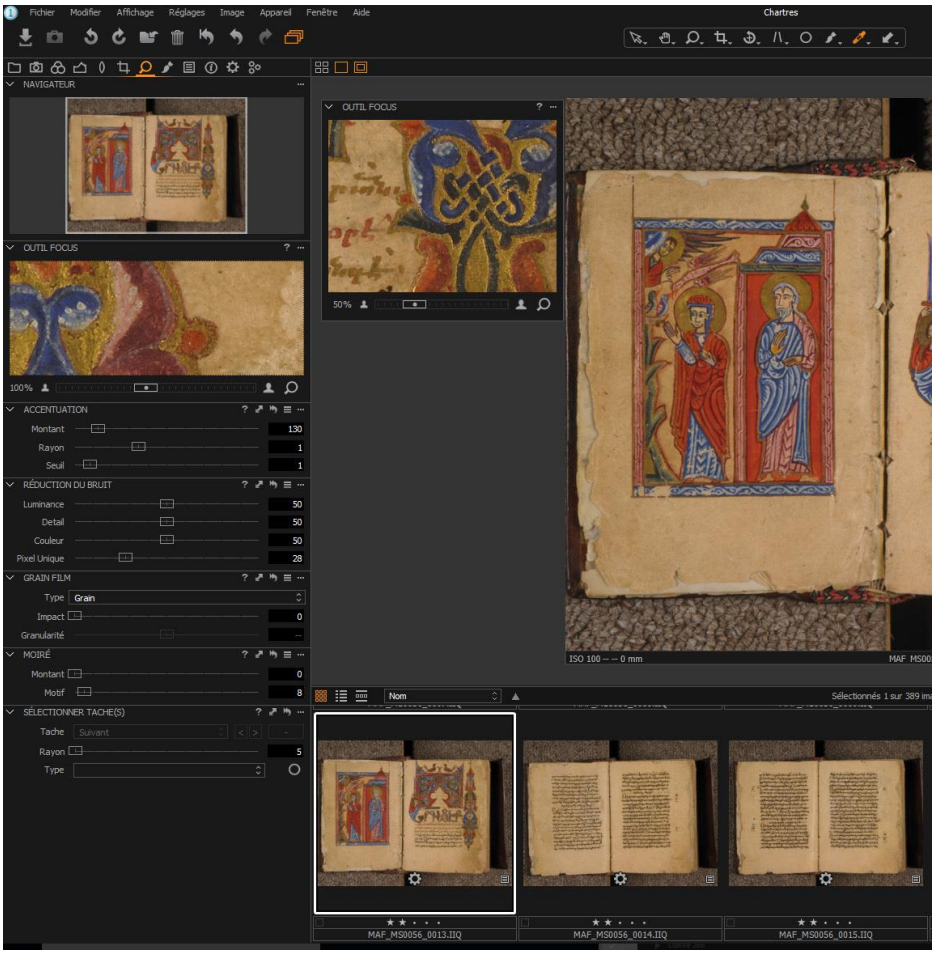


Contrôle des niveaux de l'histogramme, noir = 0 ; blanc = 255



Contrôle des vues avec l’application Capture One Pro

Le photographe procède également au contrôle permanent de la netteté de l’image, il dispose d’un outil « Focus » lui permettant de voir à la taille réelle des pixels si la mise au point est correcte.



11. Formats de fichier, échantillonnage et résolution :

L'IRHT utilise des appareils moyen format professionnels comportant un capteur matriciel : dos numérique phase One et ou Hasselblad supérieur à 50 Mégapixels et conserve les images dans des formats standards.

L'IRHT conserve les données brutes de la capture numérique en :

- Raw

Le photographe procède à partir de formats Raw aux exports suivants :

- Tiff (Attention option de compression LZW désactivée)
- Jpeg (qualité maximale facteur 10)

Fichier de consultation :

L'IRHT a choisi le format JPG 2000 (jp2) comme format d'échange et de diffusion sur Internet. Les numérisations peuvent être consultées à partir de la BVMM Bibliothèque virtuelle des manuscrits médiévaux : https://bvmm.irht.cnrs.fr

- JPG 2000 (application Kakadu, viewer Mirador et serveur IIIF

Technologie IIIF (International Image Interoperability Framework) :
IIIF Server is an open source Fast CGI module written in C++ that is designed to be embedded within a host web server such as Apache, Lighttpd, MyServer or Nginx. It supports TIFF and JPEG2000 source images.

Qualité du signal de la capture :

Les images ne doivent comporter ni bruit ni artefact liés à la capture (sous-exposition et surexposition) ou lors du traitement informatique des images notamment si les images sont accentuées à partir d'un filtre de netteté ou suite à la compression liée à l'algorithme Jpeg.

Résolution :

Le choix de la résolution est calculé en fonction de plusieurs paramètres :

- Le ratio de réduction
- La taille du document original en sortie sera au minimum de 600 DPI pour une impression papier.

Utilisation d'un capteur matriciel :

La résolution minimale est de **4056 pixels x 5368 pixels**, non interpolée.

Il est recommandé d'utiliser pour des documents de grande taille (A0) des capteurs matriciels en haute résolution supérieur à 50 Mo.
Il est recommandé de procéder à des accentuations (accentuations locales des niveaux de luminosité pour renforcer la netteté) dont la puissance sera proportionnelle à la taille du fichier.

12. Préparation des documents avant la numérisation :

Chaque document doit comporter une cote et une référence intelligible (codée) qui sera utilisée pour le signalement des fichiers.
Le référencement des vues doit se faire dès la prise de vue et il doit être compatible avec le système de catalogage des manuscrits. L'identification de la cote du document est indispensable au référencement des vues et au catalogage des documents.
La numérisation fait l'objet d'un suivi dans un tableur de type Excel qui permettra l'importation des données dans un logiciel GED ; il importe de noter le nombre fichiers reproduits et la date de la photographie.
Toutes les parties du document à numériser doivent être signalées dans un **fichier de récolement** : la reliure, les plats, les contres-plats, les gardes, les feuillets vierges, la numérotation des feuillets (folios et ou pages).
Une vue avec un réglet est également photographié sur la première vue du document.
Les feuillets sont photographiés à plat, l'appareil placé au-dessus du document à 180° de façon à éviter tout type de déformation.

Remarque : Les documents devront être contrôlés avant d'être numérisés et, si nécessaire remis en ordre et foliotés.

13. L'organisation des fichiers numériques

L'arborescence informatique sur lequel sont stockés les fichiers provenant de la numérisation doit être explicite :

- Pays / Ville / Etablissement / Cote

Il faut attribuer un nom à chaque fichier numérique. Le nommage est une opération délicate et lourde de conséquences sur l'exploitation ultérieure du fichier. Il faut établir un plan de nommage exploitable lors des campagnes de numérisation ultérieures. La structure du nommage doit être élaborée en envisageant tous les cas de figure pour l'ensemble des collections.
Il est préférable de ne pas inscrire le nom de l'ouvrage numérisé dans le nom du fichier, mieux vaut privilégier sa cote. Le nommage des fichiers utilisé par l'IRHT respecte la norme ISO 9660 niveau 2* qui est décrite dans le guide de la Direction des Archives Nationales de France afin d'assurer une totale compatibilité avec les systèmes informatiques. L'identifiant ne doit pas comporter de caractères spéciaux.

Identification des fichiers :

L'IRHT utilise deux systèmes de nommage pour signaler les fichiers images : le format RCR et le format ARK* qui est préconisé par la BNF et le Cines notamment pour la diffusion et pour l'archivage pérenne.

Format RCR (bi-feuillet)

Plan de nommage alphanumérique codé mis en place par le Ministère de la Culture **et** la MRT, les recommandations se trouvent sur le site du CCfr. **Catalogue collectif de France.** http://www.culture.gouv.fr/mrt/numerisation/fr/gestion_fonds_images/idprodbi.htm

Remarque et information :

C'est l'ABES qui est chargée d'attribuer le numéro RCR, identifiant national, (pour les bibliothèques ou centres de documentations publics ou privés qui n'en ont pas encore, et c'est le CCFR qui crée l'identifiant pour les bibliothèques qui n'appartiennent pas au réseau SUDOC ou SUDOC-PS (ancien CCN, catalogue collectif nationale de publications en série), la notice du Répertoire des centres de ressources.

Ce numéro est formé de plusieurs séquences :

- Les 2 chiffres du département
- Les 3 chiffres de la commune
- Un code à deux chiffres indiquant le type de bibliothèque : (51 pour les archives, 62 bibliothèque municipale, etc.…)
- Un numéro à 2 chiffres, séquentiel, allant de 01 à 99, qu'on attribue après s'être assuré qu'il n'en existe pas déjà dans notre fichier pour un autre organisme.

*ARK : Archival Resource Key (ARK) est un système d'identifiants basé sur la norme URI assurant opacité, extensibilité et indépendance, c'est-à-dire les critères nécessaires pour garantir l'identification d'une ressource sur le long terme. [http://NMA/]ark:/NAAN/Name[Qualifier]

Le nom du fichier se présente ainsi :

Exemple pour le Musée de Cluny on obtient le nom de fichier suivant :

75105510_MS0254_0001.Tif

Explications

75105510_MS0254_0001.Tif

En bleu le code « Identifiant des organismes producteurs (IDPROD) »

En vert la **cote CGM et ou numéro d'inventaire** propre à l'établissement détenteur ou dépositaire du document

En orange = une suite alphanumérique incrémentée automatiquement pour chaque nouvelle vue par l'appareil de prise de vue.

En violet : l'extension du format de fichier.

Format RCR (mono-page adapté pour la visionneuse mirador)

Dans le cas d'une prise de vue en mode page à page il convient de distinguer les feuillets recto et les feuillets verso notamment pour bénéficier dans la visionneuse web Mirador (Technologie IIIF) de la fonction bookView.

L'IRHT utilise le suffixe A pour désigner le recto du feuillet et le suffixe B pour désigner le verso du feuillet.

Exemple avec la bibliothèque d'Arras :

620416201_MS0034_0844_0001A
620416201_MS0034_0844_0001B
620416201_MS0034_0844_0002A
620416201_MS0034_0844_0002B

14. Métadonnées

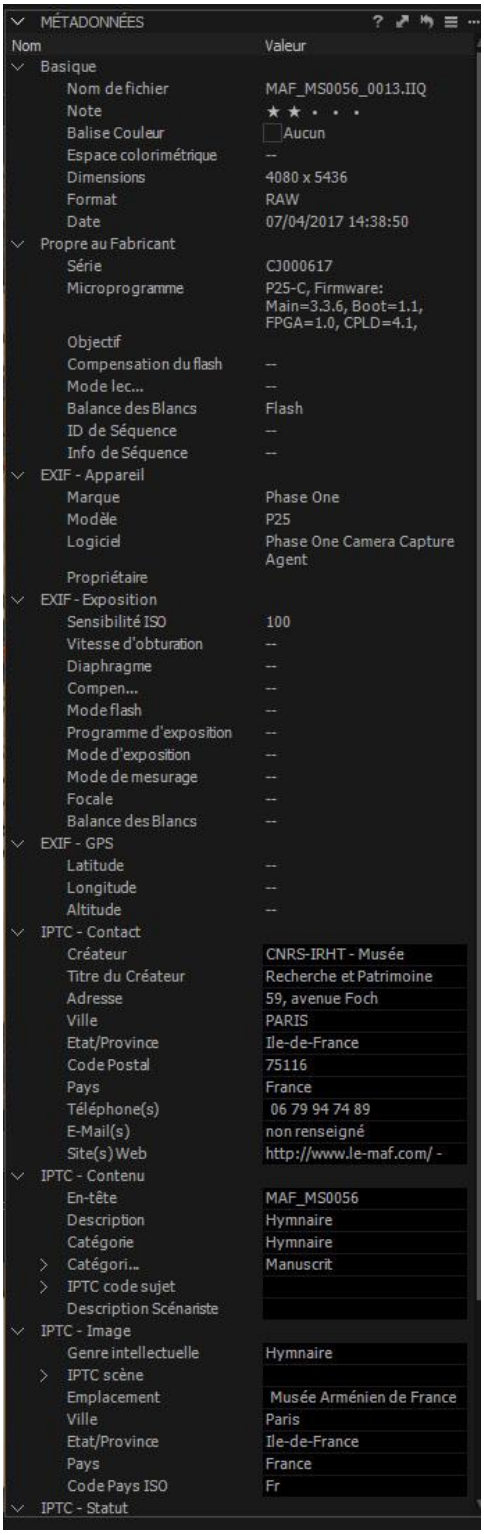
Le modèle de métadonnées permet la description du contenu du document et permet d'enregistrer les caractéristiques techniques de l'acquisition.

Ces métadonnées sont balisées dans un format d'encodage respectant la norme IPTC, Dublin Core et XML.

Le système d'acquisition de l'IRHT inclut une identification algébrique, dans le but de conserver en file système l'ordre et la structure du document original, liste des pages, liste des folios ou des vues.

Chaque fichier numérique comporte des métadonnées. Pour renseigner ces métadonnées dans les images le photographe utilise un modèle de métadonnée qui est encapsulé dans l'en-tête des fichiers Raw, Tiff, JPG et JPG 2000 pour chaque manuscrit.

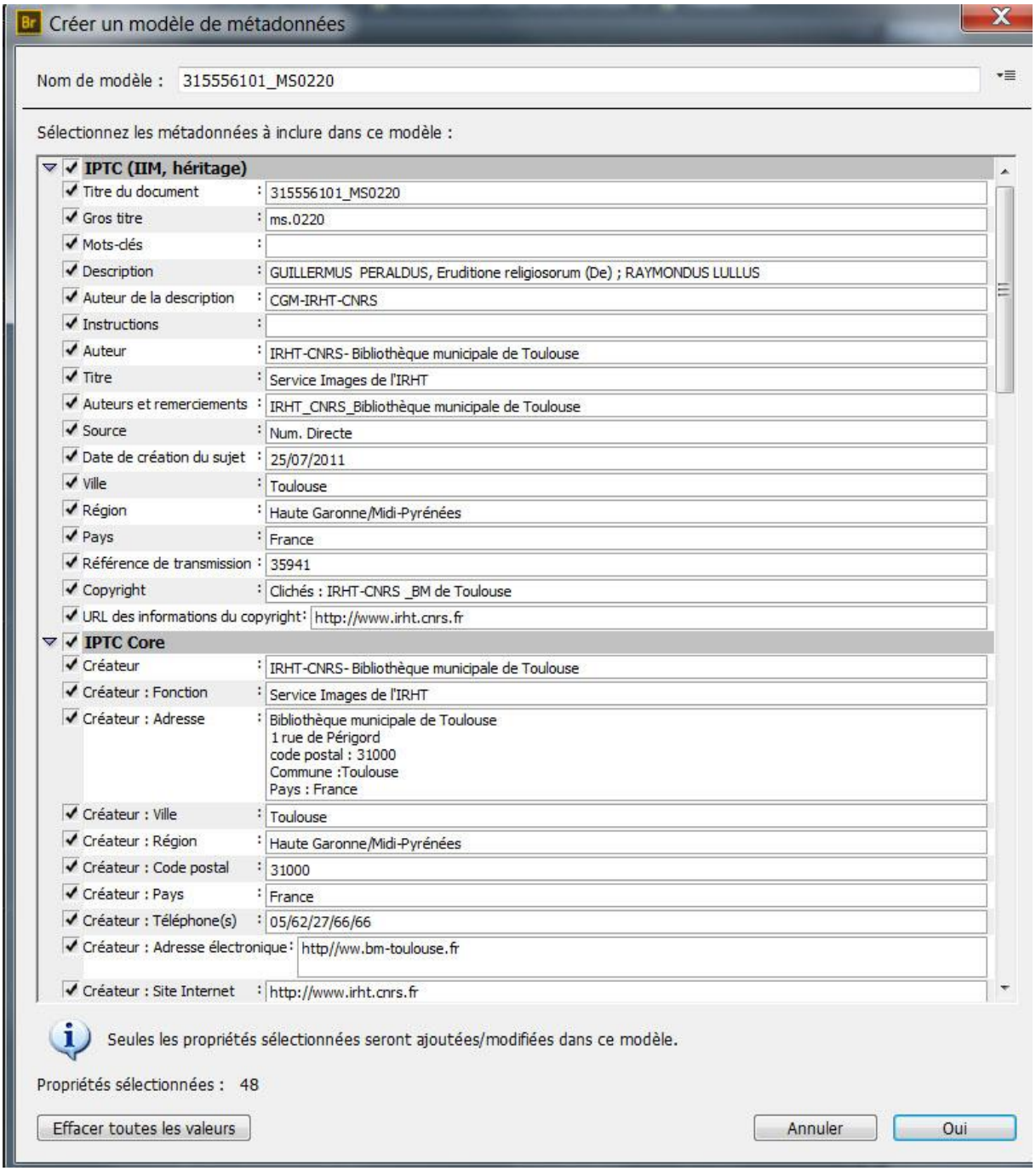
Exemple ci-dessous d'un modèle utilisé pour la numérisation des manuscrits du musée Arménien de France et à droite d'un modèle utilisé lors de la campagne de l'ENSBA.



MÉTADONNÉES	
Nom	Valeur
Basique	
Nom de fichier	MAF_MS0056_0013.IIQ
Note	★ ★ • • •
Balise Couleur	Aucun
Espace colorimétrique	--
Dimensions	4080 x 5436
Format	RAW
Date	07/04/2017 14:38:50
Propre au Fabricant	
Série	CJ000617
Microprogramme	P25-C, Firmware: Main=3.3.6, Boot=1.1, FPGA=1.0, CPLD=4.1,
Objectif	
Compensation du flash	--
Mode lec...	--
Balance des Blancs	Flash
ID de Séquence	--
Info de Séquence	--
EXIF - Appareil	
Marque	Phase One
Modèle	P25
Logiciel	Phase One Camera Capture Agent
Propriétaire	
EXIF - Exposition	
Sensibilité ISO	100
Vitesse d'obturation	--
Diaphragme	--
Compen...	--
Mode flash	--
Programme d'exposition	--
Mode d'exposition	--
Mode de mesurage	--
Focale	--
Balance des Blancs	--
EXIF - GPS	
Latitude	--
Longitude	--
Altitude	--
IPTC - Contact	
Créateur	CNRS-IRHT - Musée
Titre du Créateur	Recherche et Patrimoine
Adresse	59, avenue Foch
Ville	PARIS
Etat/Province	Ile-de-France
Code Postal	75116
Pays	France
Téléphone(s)	06 79 94 74 89
E-Mail(s)	non renseigné
Site(s) Web	http://www.le-maf.com/ -
IPTC - Contenu	
En-tête	MAF_MS0056
Description	Hymnaire
Catégorie	Hymnaire
Catégori...	Manuscrit
IPTC code sujet	
Description Scénariste	
IPTC - Image	
Genre intellectuelle	Hymnaire
IPTC scène	
Emplacement	Musée Arménien de France
Ville	Paris
Etat/Province	Ile-de-France
Pays	France
Code Pays ISO	Fr
IPTC - Statut	

Informations IPTC	
Vue:	Grand
Contact	
Créateur:	IRHT-CNRS et ENSBA de Paris
Titre du Travail:	751062305_PC_22788
Adresse:	14, rue Bonaparte
Ville:	Paris
Etat:	Bibliothèque de l'Ecole des Beaux-Arts
Code Postal:	75272
Pays:	France
Téléphone:	01.47.03.54.34
Email:	juliette.jestaz@beauxartsparis.fr
Site Internet:	http://www.irht.cnrs.fr
Contenu	
Gros Titre:	751062305_PC_22788
Description:	Bible d'Admont (fragment)
Mots Clés:	Bible d'Admont (fragment)
Code du Sujet:	
Description de l'Editeur:	
Image	
Date Crée:	2011-11-30 17:12:23
Genre:	
Tableau IPTC:	
Emplacement:	
Ville:	Paris
Etat:	Bibliothèque de l'Ecole des Beaux-Arts
Pays:	France
Code du Pays:	FR
Statut	
Titre:	751062305_PC_22788
Identifiant de la Tâche:	100045112
Instructions:	100045112
Fournisseur:	IRHT-CNRS
Source:	num. directe
Notice des...s d'Auteur:	Clichés : IRHT-CNRS _ENSBA de Paris

Métadonnées dans « Capture One ». Métadonnées dans « Phocus ».



Contrôle du modèle de métadonnées XMP à partir de Photoshop CS

15. Archivage pérenne :

L'IRHT utilise le JPG 2000 qui est beaucoup plus léger > gain de 40 % en terme de stockage par rapport au format Tiff.
L'archivage des fichiers est assuré par la plateforme du Cines de Montpellier et avec le soutien logistique et la collaboration scientifique du TGIR Huma Num. Cet archivage respecte la norme OAIS (Open Archive Initiative System).

Huma-Num
la TGIR des humanités numériques

Présentation | Services et outils | Consortiums | International | Ressources

Accueil » Services et outils

Archivage à long terme

Dernière modification le 9 juillet 2015

ARCHIVER

La TGIR Huma-Num propose à la communauté des producteurs de données numériques en Sciences Humaines et Sociales un service d'archivage à long terme. Elle s'appuie, pour cette activité, sur les infrastructures et les compétences d'un centre labellisé, le Centre Informatique National de l'Enseignement Supérieur (CINES) avec qui elle a signé une convention qui vient d'être renouvelée pour une période de cinq ans (2013 – 2018).

Huma-Num accompagne les producteurs de données tout au long du processus d'archivage à long terme et est le lien entre les communautés scientifiques et le CINES. Elle procure aide et conseils aux producteurs de données pour choisir les formats de données adéquats, mais aussi pour concevoir sous une forme normalisée les données descriptives (méta-données) indispensables pour les faire entrer dans le système d'archivage du CINES.

Comment archiver vos données ?

Les données à archiver doivent présenter une valeur scientifique reconnue par la communauté dont elles proviennent. Huma-Num soumet à son conseil scientifique les demandes d'archivage à long terme.

DEVENIR UTILISATEUR

Vous souhaitez utiliser nos services ?

Ecrivez à la TGIR : cogrid@huma-num.fr

Votre demande sera examinée dans un délai d'un mois. En cas d'accord, vous recevrez un lien vous permettant de compléter un formulaire en ligne.

Consultez les conditions générales d'utilisation.

Déjà utilisateur et besoin d'assistance ?

Utilisez l'adresse : assistance@huma-num.fr

PLUS D'INFORMATION

Dossier thématique :
Les services de conservation de données

Le paquet d'information associé à l'archive SIP* prend en charge à la fois :

- les données techniques de l'image qui sont créées par l'APN ou le système de capture (données de type Exif*)
- les métadonnées métiers
- les métadonnées liées à la description matérielle et textuelle (codicologique) du document : format, dimensions, type de support, type de document, langue, datation, provenance, auteur, contenu, lieu de conservation, etc.

*SIP : Submission Information Package

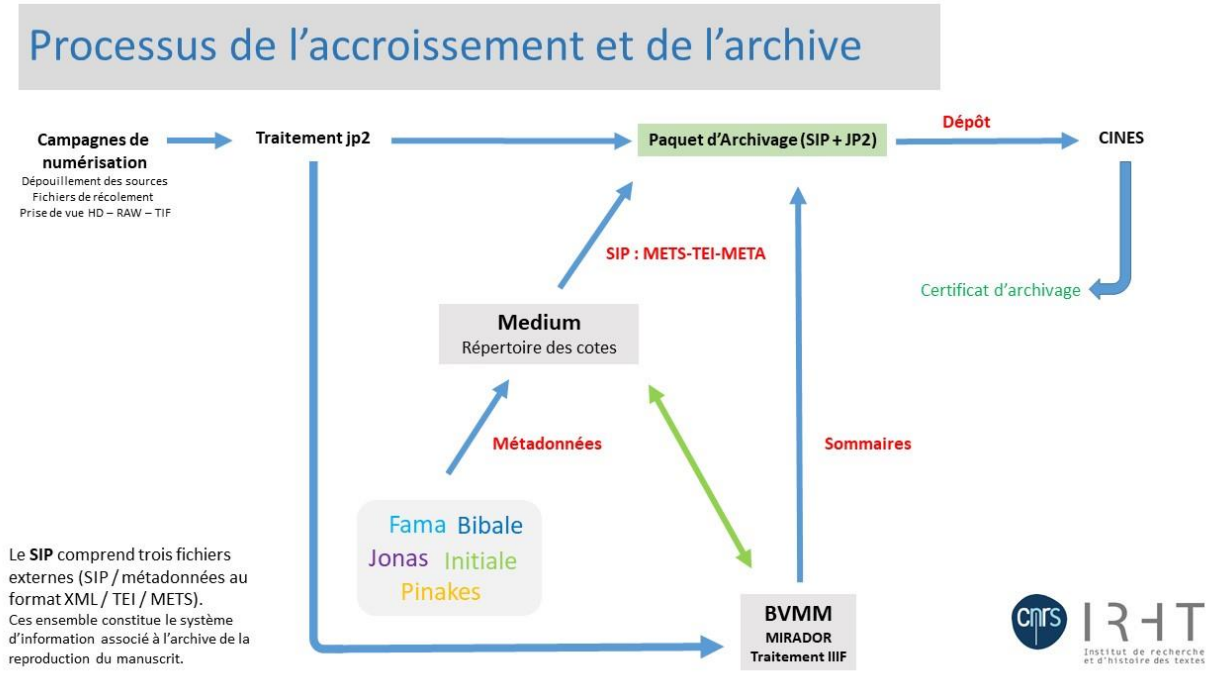
*EXIF : L'Exchangeable image file format est une spécification de format de fichier pour les images utilisées par les appareils photographiques numériques qui correspond entres autres aux propriétés techniques de la capture.

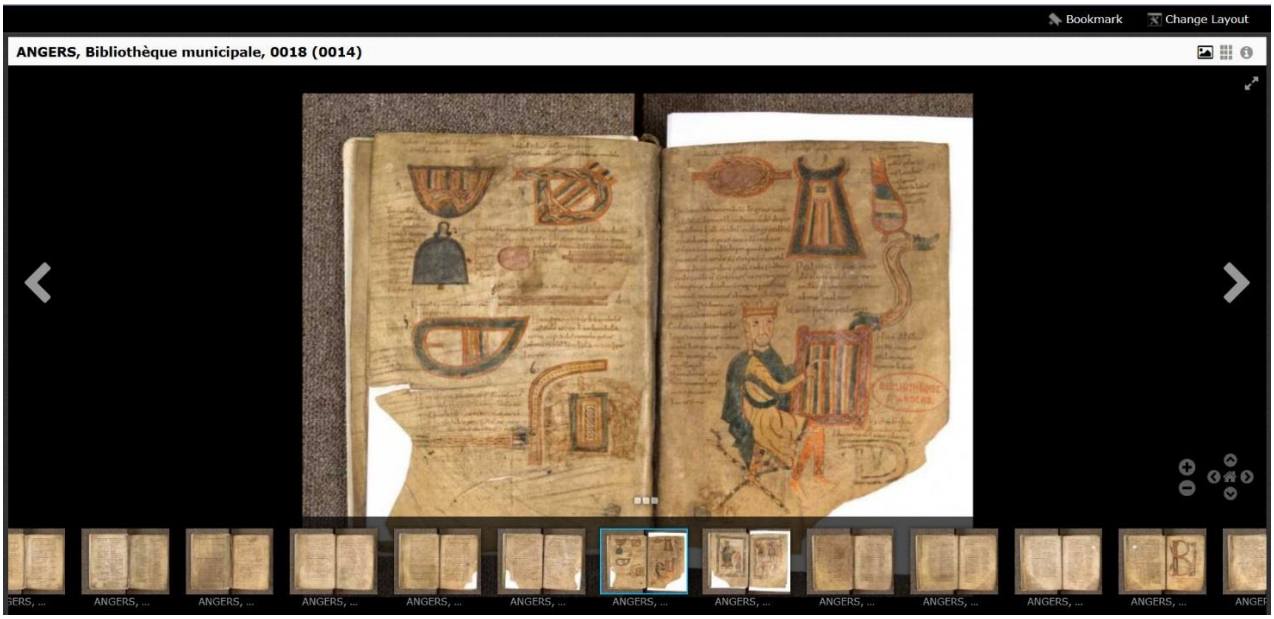
Ce format de description est décrit dans un METS, les champs respectent la norme Dublin Core et répondent dans la majorité des cas à une description minimale du document.

GED : Gestion électronique de document
Raw : format brut des images numériques.
DNG : Digital négative, format de fichier Raw d'Adobe.
JPG : Joint Expert Photographic Group
JPG 2000 : Nouvelle norme de compression du format JPG
TIF : Tagged Image File Format, format de fichier graphique bitmap.
ICC : International Color Consortium
Applications: Capture One Pro, Focus, suite Adobe (Photoshop, Bridge...), GIMP, Xn View...
IPTC : International Presse Télécommunication Council

16. Traitement et intégration des manuscrits dans la BVMM et dans la base Medium

- Sauvegarde des fichiers sur disques externes.
- Contrôle des identifiants dans la base Medium avant traitement.
- Vérification des métadonnées internes.
- Export des formats RAW en JPG , TIF.
- Conversion des fichiers Tif en jp2 à partir de Kakadu.
- Encodage des fichiers de récolements en UTF8.
- Mise en ligne des fichiers jp2 dans la BVMM.
- Vérification des liens avec la base Medium* et des identifiants ARK.
- Signalement des stockages dans la base Medium.
- Archivage des images en JPG 2000 (jp2) sur la plateforme du Cines en collaboration avec Huma-Num.





Système de visualisation Mirador utilisé pour la BVMM

*Medium : Répertoire des manuscrits reproduits et rescensés (gestion des reproductions + notices)

17. Livraison des fichiers

Le service Images procède à la livraison des numérisations au format jpeg ou jp2 et des fichiers de récolement à la fin de la campagne.