

NUMÉRISATION & QUALITÉ

GUIDE DE GESTION ET DE CONTRÔLE DE LA QUALITÉ
DANS UN PROJET DE NUMÉRISATION PATRIMONIALE

Rédaction : Chloé Brault

Contributeurs : Florence Gillet, Nicolas Roland

Relecture et direction scientifique : Anne Chardonnes

Mise en page : Chloé Brault

Comité du projet ADOCHS : Nico Wouters (CegeSoma-AGR), Frédéric Lemmers (KBR), Ann Dooms (VUB), Tan Lu (VUB/KBR), Anne Chardonnes (CegeSoma-AGR/ULB), Nicolas Roland (KBR), Florence Gillet (CegeSoma-AGR), Seth van Hooland (ULB), Chloé Brault (CegeSoma-AGR)

Centre d'Etude Guerre et Société – Studie en Documentatiecentrum Oorlog en Hedendaagse Maatschappij

Archives de l'état - 4e Direction opératoire

Square de l'Aviation 29 Luchtvaartsquare 1070 Bruxelles - Brussel

L'élaboration de ce guide n'aurait pas été possible sans le financement de la **Politique scientifique belge** dans le cadre du numéro de contrat BR/154/A6/ADOCHS.

Partenaires institutionnels



NUMÉRISATION ET QUALITÉ

Guide de gestion et de contrôle de la qualité
dans un projet de numérisation patrimoniale

Version 1.1 - Septembre 2021

Par Chloé Brault

Bruxelles
2021

TABBLE DES MATIÈRES

	À propos de ce guide	6
	Contexte de création	6
	Objet et portée	7
	Méthodologie et structure	8
01.	Numérisation : enjeux, mutations et perspectives	9
02.	La qualité : un fil rouge organisationnel	14
	2.1 Stratégie et gestion de données : le socle	16
	2.2 Le cahier des charges	18
03.	La qualité des livrables : les fichiers images	23
	3.1 Image : définition et caractéristiques	23
	La résolution et la définition	23
	La profondeur de codage	23
	Le mode	24
	L'espace colorimétrique	25
	Le format	25
	3.2 Critères d'une image de qualité	27
	La complétude	27
	L'exposition	29
	Le contraste	30
	La couleur	31
	La mise au point	31
	La résolution	32
	Le parallélisme	33
	Le flare	33
	Le vignettage	34
	Le bruit	34
	Les artéfacts	35

04.	La qualité des livrables : les métadonnées	37
	4.1 Métadonnées : définition et caractéristiques	37
	4.2 Rôle et fonction des métadonnées	39
	4.3 Prérequis à l'utilisation des métadonnées : l'interopérabilité	39
	4.4 Critères d'une métadonnée de qualité	40
	4.5 Une vision commune pour une meilleure interopérabilité	43
05.	La gestion de l'atelier de numérisation	47
	5.1 Préserver les collections patrimoniales et documentaires	47
	5.2 La gestion de la lumière	50
06.	Fiches méthodologiques pour assurer la qualité d'un projet de numérisation	53
	6.1 Définir ses référentiels	55
	6.2 Instaurer et appliquer une démarche qualité	59
	6.3 Définir la qualité des livrables	62
	6.4 Quand et comment faire intervenir le contrôle qualité	64
	6.5 Contrôler la qualité en cas d'externalisation	67
	6.6 Choisir le matériel de numérisation	69
	6.7 Choisir les éclairants	71
	6.8 Étalonner la chaîne de numérisation	73
	6.9 Règles générales de prise de vue	77
	6.10 Créer et remplir le fichier de récolement des métadonnées	79
	6.11 Enregistrer et nommer les fichiers	81
	6.12 Retoucher les fichiers images	84
	6.13 Océriser les documents texte	86
	6.14 Encapsuler les métadonnées	88
	6.15 Dernier contrôle qualité	90
	6.16 Choisir son système de gestion de données numériques	91
	Bibliographie	94

À PROPOS DE CE GUIDE

Contexte de création

Depuis le milieu des années 90, les institutions culturelles sont sans aucun doute entrées dans l'ère numérique. En Belgique comme ailleurs dans le monde, gouvernements et sociétés privées multiplient les projets de numérisation à grande échelle¹ depuis le début des années 2000, et nombre d'institutions d'envergure possèdent aujourd'hui un atelier de numérisation en leur sein.

La question du contrôle de la qualité a été l'un des principaux obstacles lors de cette première phase. En effet, il est apparu que de nombreux projets avaient sous-estimé l'ampleur de cette étape, tant à un niveau humain que technique, dans le processus global de numérisation. Dans la plupart des cas, les équipes ont été confrontées à un manque de normalisation méthodologique et d'outils d'automatisation pour effectuer le travail. Elles ont donc souvent dû travailler manuellement et sans directives de procédures adaptées à leurs besoins spécifiques.

Le contrôle de la qualité est pourtant une composante essentielle à chaque étape d'un projet de numérisation si l'on veut garantir

l'intégrité et la cohérence des fichiers et des données produites, ainsi que leur conservation à long terme. Et ce tant dans le cadre de projets de numérisation externalisés que pour des projets de numérisation internes.

La multiplication d'intervenants a entraîné des différences dans le résultat final de la numérisation. Comment s'assurer que le fichier numérique réponde à un niveau attendu de qualité et qu'il soit une représentation fidèle et intégrale du document original afin que l'utilisateur ne se pose pas de questions sur l'intégrité de ce dernier ? Comment faire en sorte que des documents similaires possèdent les mêmes caractéristiques une fois numérisés, et ce, quels que soient l'institution, l'opérateur, l'atelier et même l'appareil utilisé ?

De ces réflexions est né le projet de recherche **ADOCHS** (*Auditing Digitalization Outputs in the Cultural Heritage Sector*), qui a débuté en novembre 2016.

À l'initiative des Archives de l'État - et plus spécifiquement du CEGESOMA² - en collaboration avec la Bibliothèque Royale de Belgique (KBR), la Vrije Universiteit Brussel (VUB) et l'Université libre de Bruxelles (ULB),

le projet ADOCHS a pour but d'améliorer le contrôle qualité des fichiers produits dans le cadre de la numérisation patrimoniale. Ce faisant, le projet vise à accélérer l'ensemble du processus de numérisation, à minimiser ses coûts et à augmenter la valeur des données produites dans le cadre de futurs projets.

Objet et portée

Ce guide est destiné avant tout aux services de numérisation des établissements scientifiques fédéraux belges.

Les opérateurs de numérisation y trouveront des recommandations techniques pour numériser des collections de façon adaptée et homogène. Les responsables de collections et les gestionnaires de projets y trouveront quant à eux une vue d'ensemble des processus de la numérisation, de même que les éléments articulant ces processus, afin de les aider à élaborer précisément un projet de numérisation.

Autrement dit, **ce manuel a comme objet de proposer des outils organisationnels et méthodologiques afin d'assurer la qualité des données numériques produites à chacune des étapes du processus de numérisation.**

De cette manière, chaque personne impliquée dans ce processus peut envisager son travail comme faisant partie d'une chaîne de production avec un objectif commun.

Le contrôle de la qualité est une composante essentielle à chaque étape d'un projet de numérisation si l'on veut garantir l'intégrité et la cohérence des fichiers et des données produites, ainsi que leur conservation à long terme.

L'objectif de ce guide est double :

- Standardiser les critères de contrôle de qualité
- Garantir la préservation ainsi que la (ré)utilisation à long terme du patrimoine culturel numérisé.

Ce guide tend ainsi à aboutir à une harmonisation des pratiques des institutions scientifiques et culturelles fédérales belges. De cette harmonisation découle la facilité d'échange des informations au sein de ces institutions et entre ces dernières. Les utilisateurs profiteront également de cette amélioration en voyant leur capacité de recherche augmentée et facilitée.

1 *Annoncé en 2004, le projet Google-Print, en collaboration avec les bibliothèques des universités de Harvard, de Stanford, du Michigan aux USA, de l'Université d'Oxford en Grande-Bretagne et la bibliothèque publique de New York, avait pour ambition de numériser 15 millions d'ouvrages. De cette initiative découle directement le projet Europeana et nombre de projets de numérisation d'organismes privés et publics. Dans son article, Jean-Michel Salâün retrace la naissance de ce projet dantesque et les vives réactions qu'il a suscitées à l'époque.*

Jean-Michel Salâün. *Bibliothèques numériques et Google-Print. Regards sur l'actualité : mensuel de la vie publique en France, La Documentation Française*, 2005.
https://archivesic.ccsd.cnrs.fr/sic_00001576/document

2 *Le CegeSoma, Centre d'Études et de Documentation Guerres et Sociétés contemporaines est la quatrième direction opérationnelle des Archives de l'Etat belge depuis 2016.*

Méthodologie et structure

Ce guide est divisé en six chapitres selon la structure suivante :

1 Le premier chapitre recontextualise brièvement la numérisation et ses enjeux. Sont énumérés les objectifs les plus courants de la numérisation et leurs évolutions au vu de l'informatisation toujours plus importante de nos sociétés modernes. De cette évolution découlent en effet des mutations institutionnelles s'appuyant elles-mêmes sur l'évolution des attentes et besoins des utilisateurs des services d'archives et bibliothèques.

2 Le deuxième chapitre définit la notion de qualité selon la norme internationale ISO-9001³. Cette définition, généraliste, est ensuite adaptée et affinée aux spécificités du processus de la numérisation. L'élaboration stratégique et intellectuelle de tout projet est ici abordée à travers trois documents fondamentaux : la stratégie de gestion de données, la politique de numérisation et enfin le cahier des charges.

3 Les chapitres trois et quatre s'intéressent à la qualité des livrables, et plus particulièrement aux images et aux métadonnées qui y sont associées. Ces deux notions sont tour à tour explicitées et caractérisées pour pouvoir identifier les critères leur permettant d'être définies comme de bonne qualité.

5 Le cinquième chapitre aborde quant à lui l'environnement de la numérisation et la bonne gestion de l'atelier. La numérisation patrimoniale a en effet la particularité de traiter des objets précieux — qu'ils soient modernes ou anciens — qui nécessite des mesures particulières pour éviter leur dégradation lors de leur manipulation.

6 Enfin, le sixième et dernier chapitre propose une série de fiches récapitulatives pour garantir la qualité d'un projet de numérisation. Ces fiches rassemblent des notions-clé pour chaque étape de contrôle qualité et suivent le déroulé chronologique de la chaîne de la numérisation. Chaque fiche comprend ainsi des recommandations pratiques, des outils et conseils ainsi que des sources complémentaires pour mener à bien la phase de contrôle décrite.

Le contenu de ce guide s'appuie sur :

- Des référentiels standardisés autour de la qualité et de la numérisation, et plus spécifiquement sur des normes ISO.
- Des référentiels méthodologiques internationaux et guides de bonnes pratiques de la numérisation.
- Des monographies généralistes autour de la numérisation, de la qualité et de son management, de l'imagerie numérique et des métadonnées.
- Les recherches menées durant le projet ADOCHS et plus particulièrement sur les travaux de Tan Lu et Anne Chardonnens sur l'amélioration de la qualité des images et des métadonnées.
- Les compétences et l'expérience des équipes de numérisation des Archives de l'État belge et de la Bibliothèque Royale de Belgique, grâce à une série d'entretiens menés au printemps 2021 auprès des opérateurs techniques et des gestionnaires de projets de numérisation.

³ Organisation Internationale de Normalisation (ISO). *Systèmes de management de la qualité – Principes essentiels et vocabulaire*, Genève : ISO, ISO 9001 :2015, 2015.

<https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v2:fr>

01.

LA NUMÉRISATION : ENJEUX, MUTATIONS, PERSPECTIVES

La numérisation est l'action de convertir une ressource analogique en une suite de codes interprétables par les machines, permettant ainsi de reproduire sous forme de documents numériques tout type de documents analogiques¹.

Derrière cette définition a priori simple se cachent en réalité de multiples enjeux impliquants les institutions initiatrices de ces projets, mais aussi leurs publics. Les chantiers de numérisation doivent ainsi concilier des impératifs d'ordre culturel, scientifique, social, économique, politique et enfin technique.

L'objectif de la numérisation des documents dépend intimement de l'exploitation prévue pour ces contenus numérisés, et sert globalement à la conservation, à la diffusion et à la valorisation des collections patrimoniales, archivistiques et documentaires.

Conservation	La numérisation réalisée à des fins de conservation consiste à reproduire le plus fidèlement possible le document original, sans chercher à l'améliorer ou à le modifier . Ce substitut permet de limiter la manipulation et donc la dégradation des collections originales.
Diffusion	La numérisation réalisée dans un objectif de diffusion vise à rendre accessible un document numérisé sur le Web ou dans une publication .
Valorisation	L'objectif de valorisation de la numérisation se manifeste par exemple à travers la création de corpus numériques facilitant l'enseignement et la recherche ou dans le cadre d'une exposition muséale.

¹ Thierry Claerr, Isabelle Westeel, *Manuel de la numérisation*, Electre, Tours, 2011, page 18.

L'avènement d'internet et ses évolutions ont partiellement conditionné l'accès à la culture et sa valorisation. En effet, le Web 2.0 associe divers systèmes sociaux complexes tels les blogs, wikis ou réseaux sociaux et envisage le Web comme un grand espace d'interactions sociales. C'est la collaboration entre les individus et les plateformes qui garantit la production continue de contenus.

En ce sens, les sites web 2.0 agissent plus comme des portails centrés sur l'utilisateur dont il faut capter l'attention plutôt que sur des sites web traditionnels. Soutenu par des algorithmes de recommandation, le Web 2.0 fonctionne sur le principe de similarité entre les profils et les contenus. Ces similarités sont notamment définies grâce aux données d'usage des utilisateurs, ce qui permet de créer des groupes aux caractéristiques relativement homogènes par et vers lesquels diffuser l'information.

Dans ce contexte, les métadonnées – qui structurent l'identification et la description des contenus numériques – sont évidemment centrales pour rejoindre ce cercle de diffusion et trouver une audience. L'importance accordée à la manière de décrire un contenu ou un site web pour optimiser son référencement s'est accrue et la question de la structuration de l'information et de sa lisibilité pour les machines est devenue une problématique centrale dans notre façon d'aborder le Web.

C'est tout l'enjeu du projet du Web 3.0, plus communément désigné sous le terme de Web sémantique, dont l'objectif n'est plus de construire des plateformes de consultation de données, mais bien de rendre accessible et lisible de l'information de référence aussi bien pour les machines que pour les humains.

Cet objectif implique entre autre que le format des métadonnées soit suffisamment souple pour permettre la consultation et la diffusion du contenu numérique décrit, tout en favorisant le croisement et la transformation de l'ensemble des contenus présents sur le web.

Ce projet prend tout son sens lorsque l'on pense à la diffusion numérique des collections patrimoniales, archivistiques et documentaires : une telle technologie permettrait par exemple **de rassembler les informations connues autour de corpus souvent dispersés à travers plusieurs institutions tant au niveau national qu'international.**

Dans ce contexte, les métadonnées sont centrales pour trouver une audience.

Cette démarche demande cependant aux acteurs du secteur culturel et scientifique une grande cohésion de même qu'une stratégie commune. Les métadonnées accompagnant les contenus numériques doivent en effet être suffisamment bien renseignées pour que ces derniers puissent être repérables² par leur public potentiel.

C'est avec cet objectif que la Commission européenne a récemment revu le plan de développement³ de son projet phare pour la

2 *La repérabilité d'un contenu numérique correspond à la facilité ou non pour un contenu en ligne d'être trouvé par les internautes, notamment par ceux qui ne cherchaient pas précisément le contenu en question. Ce potentiel dépend de divers facteurs, notamment les métadonnées et les mots-clés qui lui sont associés, son référencement, la publicité dont il fait l'objet, les algorithmes du moteur de recherche utilisé ainsi que les habitudes de navigation des internautes.*

Pour plus de détails sur le sujet consultez - Josée Plamondon. Êtes-vous repérables ? Documenter vos contenus pour qu'ils soient compris par des moteurs de recherche, 2019.
http://ipf.ca/wp-content/uploads/sites/2/2019/05/guide-documentation-webseries_20171128_vf-maj_20191405.pdf

valorisation du patrimoine, **Europeana**⁴. Europeana est une plateforme numérique lancée en novembre 2008 donnant accès aux ressources numériques des institutions culturelles des États membres, et compte aujourd'hui près de 3.500 institutions participantes. À noter qu'Europeana n'archive pas les œuvres, mais sert uniquement d'interface de recherche. Les institutions participantes s'engagent donc à numériser leurs contenus, à les rendre accessibles sur le Web et à assurer leur conservation sous forme numérique.

L'enjeu du document stratégique proposé par la Commission est ainsi d'**offrir une ligne directrice commune concernant les problématiques de la numérisation, et notamment d'encourager les institutions participantes à s'approprier la question des métadonnées en vue de décupler leur visibilité, mais aussi le potentiel de diffusion de leurs collections sur le Web.** Les différentes mesures proposent ainsi d'œuvrer à une meilleure transition digitale en travaillant sur les points suivants :

- ***L'amélioration de la qualité des métadonnées et de leur interopérabilité.***
- ***L'usage nouvelles technologies telles que les algorithmes d'apprentissage automatique pour enrichir automatiquement ou semi-automatiquement les enregistrements de métadonnées d'une manière plus rapide et plus évolutive.***
- ***Le renforcement des infrastructure nationales et le développement d'interconnexion interinstitutionnelle plus forte.***

- ***L'amélioration des compétences techniques des acteurs du secteur par la mise en place d'un centre de connaissances offrant aux professionnels du patrimoine culturel des lignes directrices sur des sujets tels que la numérisation, l'enrichissement des métadonnées, l'interopérabilité sémantique, la création de contenu, l'octroi de licences, la réutilisation, les modèles commerciaux et l'innovation.***

Cette perspective d'évolution quant au cadre de diffusion des contenus numériques demande néanmoins des ajustements pour les professionnels du patrimoine, des archives et des bibliothèques.

Passer d'un accès restreint et individuel à une ouverture sur un corpus de masse a demandé, et demande toujours aux institutions, une restructuration en profondeur de leurs politiques de diffusion et des moyens soutenant cette politique. Ce changement de paradigme a également modifié le statut et les compétences demandées aux acteurs du secteur : de conservateurs du patrimoine public, les bibliothécaires, archivistes et muséologues sont devenus des fournisseurs de contenus culturels⁵. Cette mutation implique le développement d'une nouvelle expertise technique et demande de porter une attention particulière à l'évolution des attentes et besoins des utilisateurs de ces services.

3 *Europeana, Strategy 2020-2025 Empowering Digital Change, mai 2020, https://pro.europeana.eu/files/Europeana_Professional/Publications/EU2020StrategyDigital_May2020.pdf*

4 *Europeana Pro, Présentation, n.d., <https://pro.europeana.eu/about-us/mission>*

5 *Annabelle Boutet et Karine Roudaut, « Les enjeux de la numérisation et de l'ouverture d'archives : le point de vue des professionnels », Terminal <http://journals.openedition.org/terminal/1220>*

Une telle étude a été menée en Belgique entre 2015 et 2017 dans le cadre du projet de recherche **MADDLAIN**⁶. Son objectif n'était pas seulement de fournir des données sur les pratiques et les besoins des utilisateurs en matière d'accès à l'information numérique, mais aussi de faire progresser les institutions dans la gestion des méthodes et des outils qui les aideront à acquérir une connaissance plus approfondie de leurs publics.

Cette perspective d'évolution demande des ajustements pour les professionnels du patrimoine, des archives et des bibliothèques.

De cette étude⁷ ressortent deux termes fondamentaux : le besoin d'**accessibilité** et de **communication**. Le rapport final du projet stipule à cet égard que « *les principaux points d'insatisfaction de l'enquête menée soulignent en effet les difficultés pour les utilisateurs inexpérimentés à trouver l'information recherchée, les lacunes dans les informations présentées – description inadéquate ou incomplète des documents – ainsi que le manque de communication concernant les documents qui sont inscrits ou non dans le catalogue numérique. Parallèlement, la majorité des utilisateurs comprend que l'ensemble des collections ne puissent pas être numérisé et disponible en ligne. Ils demandent cependant des explications sur les procédures techniques sous-jacentes, sur les choix effectués en matière de numérisation et sur les restrictions potentielles* ».

Ces conclusions mettent en lumière toute l'importance d'aborder la numérisation des collections patrimoniales, archivistiques et documentaire à travers la notion de **qualité**. **Qualité des livrables bien sûr, mais aussi qualité des processus et notamment de la phase d'élaboration intellectuelle et stratégique du projet**. Ces deux pendants de la qualité appréhendés ensemble, il devient possible de mieux maîtriser l'ensemble des processus de la numérisation et, *in fine*, de proposer des contenus numériques fiables et pertinents, comme proposé dans les chapitres suivants.

6 Né d'une collaboration entre la Bibliothèque Royale de Belgique, les Archives de l'État en Belgique et le Centre d'Études Guerre et Société (Cegesoma), le projet **MADDLAIN** (2015–2017) avait pour objectif d'analyser le comportement et les besoins des différents publics des institutions du pôle documentation de la Politique scientifique fédérale en terme d'accès à l'information numérique.

7 Florence Gillet, Jill Hungenaert, Melissa Hodza, et al., *Identifying Needs to Modernize Access to Digital Data in Libraries and Archives (MADDLAIN) : final report*, Brussels : Belgian scientific Policy, 2018. http://www.belspo.be/belspo/brain-be/projects/FinalReports/MADDLAIN_%20final%20report.pdf

POUR ALLER PLUS LOIN

FONDS INDÉPENDANT DE PRODUCTION, *Êtes-vous repérables? Guide pratique pour documenter vos contenus*, novembre 2017, [En ligne],
<http://ipf.ca/FIP/ressources/etes-vous-reperables/#1512498362849-3b320f4f-a27f>

CHAN, L., MOUNIER, P., Connecting the Knowledge Commons – From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers. Marseille : OpenEdition Press, 2019.

CANDELA, G., ESCOBAR, P., CARRASCO, R.-C., MARCO-SUCH, M., « A linked open data framework to enhance the discoverability and impact of culture heritage », *In: Journal of Information Science*, 45(6), 2019, pp. 756-766.

FREIRE, N., VOORBURG, R., CORNELISSEN, R., DE VALK, S., MEIJERS, E., ISAAC, A. « Aggregation of Linked Data in the Cultural Heritage Domain: A Case Study in the Europeana Network », *In: Information*, 10(8), 2019.

JONES, E.; SEIKEL, M., Linked Data for Cultural Heritage; Facet Publishing: Cambridge, UK, 2016.

ACCART, J.-P., Le Métier de Documentaliste. Avec la collaboration de Réthy Marie-Pierre. Éditions du Cercle de la Librairie, 2015.

THIAULT, F., « Mutations des métiers de l'information-documentation : vers l'émergence d'une culture de l'information numérique », *In: Les Cahiers de la SFSIC*, Société française des sciences de l'information et de la communication, 2012, pp. 59-62, [En ligne],
<https://hal.univ-lille.fr/hal-00955279/document>

GILLET, F., HUNGENAERT, J., HODZA, M., et al., Identifying Needs to Modernize Access to Digital Data in Libraries and Archives (MADDLAIN) : final report. Bruxelles : Police Scientifique belge, 2018, [En ligne],
http://www.belspo.be/belspo/brain-be/projects/FinalReports/MADDLAIN_%20final%20report.pdf

02.

LA QUALITÉ : UN FIL ROUGE ORGANISATIONNEL

Définie par la norme ISO 9000 comme « l'ensemble des caractéristiques et des propriétés d'un produit, d'un processus ou d'un service qui influent sur son aptitude à répondre à des besoins identifiés ou implicites¹ », la qualité est une notion large.

Tous les projets peuvent en théorie s'approprier la notion de qualité ainsi que les normes internationales, et ce, quels que soient, les objectifs visés, les technologies employées, les ressources utilisées ou encore la nature de l'organisme.

Les bibliothèques, services d'archives et musées s'appuient d'ailleurs déjà en partie sur ces recommandations dans leurs activités, notamment dans le cadre de projets de numérisation. Le contrôle qualité, étape à part entière du processus, est ainsi le plus souvent en bout de chaîne et porte sur le contrôle des livrables : plus précisément sur l'évaluation de la qualité des images, des métadonnées et de l'intégrité des fichiers. En somme, la qualité revient ici à garantir l'adéquation entre les objectifs initiaux du projet et les résultats

obtenus. Mais assurer la qualité d'un projet de numérisation ne saurait être réduit au seul contrôle a posteriori des livrables : en cas d'erreurs, les frais engendrés sont trop importants et la marge de manœuvre pour rectifier les erreurs trop faible. Si cette étape de vérification finale est évidemment essentielle au processus de numérisation, elle doit donc s'accompagner d'un travail de planification et de contrôle tout au long de la chaîne de la numérisation.

En ce sens, la notion de qualité doit être perçue non pas comme une étape spécifique, mais comme un outil de management de projet guidant l'ensemble des étapes, un fil rouge garantissant la maîtrise de l'ensemble des processus.

¹ Organisation Internationale de Normalisation (ISO), Systèmes de management de la qualité – Principes essentiels et vocabulaire. Genève : ISO, ISO 9001 :2015, 2015.

Sur la base de cette définition, la notion de qualité implique dès lors :

- Des compétences techniques et une connaissance approfondie des images et des métadonnées, de même qu'une bonne compréhension du matériel de numérisation.
- Une vue d'ensemble des problèmes de qualité les plus fréquents et de leurs causes, afin de développer une méthodologie cohérente en amont du projet et d'appréhender au mieux les imprévus durant le processus de numérisation.
- La gestion de l'environnement de travail, qui comprend l'organisation des espaces, la maîtrise de la lumière, ainsi que la mesure et l'analyse des conditions atmosphériques, pour respectivement fluidifier la chaîne de travail, garantir les meilleures conditions pour la prise de vue et minimiser l'altération potentielle des collections patrimoniales.

La définition et l'utilisation d'outils

- organisationnels pour articuler les différents postes de la chaîne de numérisation et faciliter la communication entre les opérateurs de l'atelier.

Plus largement, chaque projet de

- numérisation devrait d'inscrire au sein d'une politique globale de conservation, de numérisation, et de diffusion de ses contenus numérisés, de même que dans une stratégie de gestion des données (Digital Data Strategy²).

En somme, envisager la notion de qualité comme un moteur revient à « pérenniser une action, optimiser les processus de fabrication, améliorer l'efficacité du fonctionnement d'une organisation, améliorer la qualité de ses produits et services, innover et partager de la valeur ajoutée [et enfin] rendre compte d'un usage maîtrisé de fonds investis pour la satisfaction des utilisateurs finaux³ ».

La notion de qualité doit être perçue non pas comme une étape spécifique, mais comme un outil de management de projet guidant l'ensemble des étapes, un fil rouge garantissant la maîtrise de l'ensemble des processus.

Autrement dit, il ne s'agit plus donc plus uniquement de constater la qualité d'une production a posteriori, mais de déterminer les critères de qualité d'un produit numérique, d'identifier les procédures, les enchaînements et les interactions entre les différentes étapes de la chaîne de numérisation, d'élaborer une qualité de services et enfin d'anticiper les attentes et besoins des utilisateurs afin de proposer un contenu qualitatif et pertinent.

2 Valentina Bachi, Antonella Fresa, Claudia Pierotti, Claudio Prandoni, The digitization age: mass culture is quality culture. challenges for cultural heritage and society. In: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarić, R., Yen, A.-Y., Quak, E. (eds.) EuroMed 2014, 17 p. et Mike Fleckenstein, Lorraine Fellows, Data Quality. In: Modern Data Strategy. Springer, Cham, 2018.

3 Thierry Claerr, Isabelle Westeel, Manuel de la numérisation, Electre, Tours, 2011, page 269.

Bien documenté, chaque projet de numérisation peut ainsi être évalué et permettre l'amélioration de l'ensemble des protocoles et processus, inscrivant l'action de l'organisation de manière durable dans le temps. À noter que cette approche doit évidemment être pensée selon les moyens techniques, humains et financiers de l'institution. **Pour être opérationnel, le système doit être adapté aux ressources réelles de l'organisation initiatrice du projet.**

2.1 Stratégie de gestion de données et politique de numérisation : le socle

Les circuits de décision opérationnels et stratégiques sont essentiels au succès d'un projet, et ce quels que soient le nombre de documents à numériser ou la nature du chantier.

Il importe de bien dissocier le niveau opérationnel – qui comprend l'ensemble des règles de traitement d'un document particulier et découle de la politique de numérisation de l'établissement – des questions véritablement stratégiques qui relèvent des politiques de gestion et de développement des institutions culturelles et scientifiques, ainsi que des stratégies de gestion des données et de leur qualité.

Bien que ce guide n'ait nullement pour objectif de développer en détail l'élaboration d'une stratégie de gestion de données numériques (Digital Data Strategy) ou d'une politique de numérisation, il est essentiel de souligner leur importance et leurs différences, car certaines questions opérationnelles ne peuvent être résolues sans un arbitrage de niveau stratégique :

La stratégie de gestion de données regroupe toutes les pratiques de management des données, c'est-à-dire du patrimoine informationnel en tant que ressource de valeur au sein d'un organisme, d'une institution ou d'une entreprise. Son objectif est donc de **définir, approuver et communiquer les stratégies, politiques, normes, architectures et procédures entourant la gestion des données et assurer le suivi et respect de ces dernières.** Plus les données créées, classées et traitées par l'organisme sont de bonne qualité, plus le système d'information est lisible, pertinent et fiable, œuvrant ainsi à l'amélioration globale de l'efficacité de l'organisme.

La politique de numérisation découle directement de la stratégie de gestion de données, mais a comme périmètre plus restreint les projets de numérisation de l'organisme. **Ce document regroupe quant à lui l'ensemble des principes administrant la numérisation des collections patrimoniales, documentaires et archivistiques. Il permet ainsi de définir les enjeux et objectifs de l'institution en matière de sélection des documents à numériser, de leur conservation, de leur gestion et de leur diffusion sous forme numérique.** Il comporte également l'ensemble des modalités d'élaboration et de réalisation des projets ainsi que le rôle des équipes durant le processus.

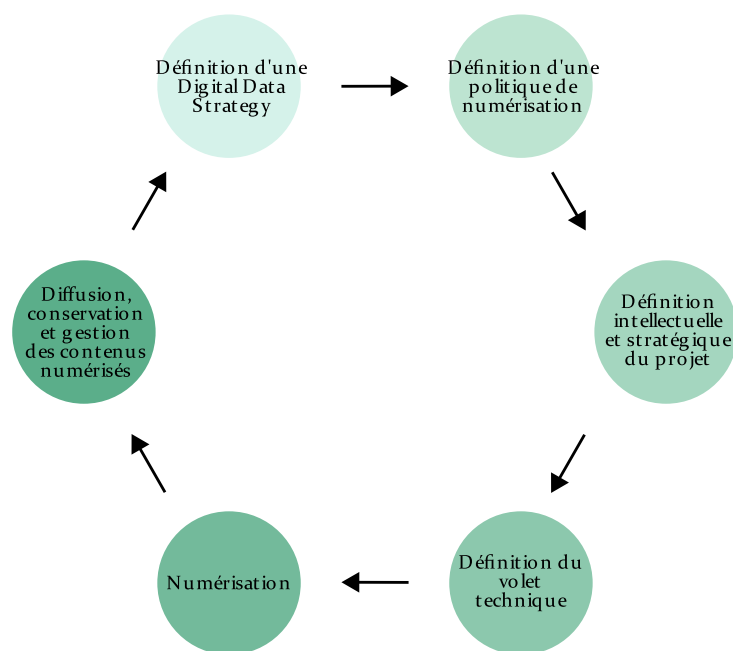


fig. 1 Le cycle complet de la numérisation

Le cadre institutionnel que représente la *Digital Data Strategy* et la politique de numérisation facilitent la définition stratégique et intellectuelle des projets de numérisation, notamment en ce qui concerne les objectifs de ces chantiers (conservation / diffusion / valorisation) et leur portée sur le long terme.

Il importe **qu'un organe de décision légitime et officiel soit identifié dès le lancement du projet** de numérisation pour pouvoir être saisi rapidement de ces questions en cas de litige avec un prestataire externe par exemple.

La **définition stratégique et intellectuelle du projet de numérisation** consiste essentiellement en l'élaboration du **cahier des charges, qui trace le cadre méthodologique du projet de numérisation** et permet à son tour de définir le volet technique du projet (choix du matériel, des logiciels, des règles de prise de vue, du transport éventuel, etc.).

Une fois les contenus numérisés, ils sont sauvegardés et stockés sur un système de gestion électronique des documents (GED) ou un système d'archivage électronique (SAE) en vue de leur diffusion et/ou de leur conservation... dont la mise en place, les objectifs et la gestion répondent directement à la stratégie de gestion de données. Chaque livrable issu d'un projet de numérisation trouve ainsi sa place dans un cercle beaucoup plus large dont il convient d'identifier et d'analyser tous les aspects.

2.2 Le cahier des charges

La gestion de la qualité d'un projet de numérisation s'élabore dès la rédaction du cahier des charges.

Comme mentionné précédemment, ce document s'inscrit dans une stratégie plus large de l'organisation commanditaire et découle directement des objectifs de l'institution en matière de conservation et de diffusion de ses collections patrimoniales et documentaires.

Le cahier des charges doit évidemment prendre en compte la faisabilité juridique du projet en s'appuyant sur les lois qui régissent la propriété intellectuelle. Enfin, il doit refléter les ressources matérielles ainsi que les choix organisationnels de l'organisme initiateur du projet : l'institution possède-t-elle un atelier de numérisation ? Quel est son rôle ? Une partie ou l'entièreté du processus doit-il être externalisé ? L'équipe possède-t-elle les compétences et le matériel nécessaire pour toutes les étapes ? Quels sont les objectifs du projet de numérisation et quelle exploitation est prévue pour les contenus numérisés ?

La réponse à chacune de ces questions influence les prestations à effectuer, la qualité attendue, le seuil de qualité minimale ou encore la gestion du contrôle qualité.

En vue de mieux maîtriser la qualité d'un projet de numérisation, voici ce qu'un cahier des charges doit comprendre :

01 Une description précise de l'institution et de ses objectifs

Décrire la nature et l'objectif général du projet, les types de documents concernés, le résultat attendu, l'exploitation prévue. En somme, il s'agit d'un résumé de la politique de numérisation de l'institution, de ses moyens et de ses ambitions concernant le projet.

02 Un résumé des prestations à réaliser

Récapitulatif des étapes de réalisation du projet. En cas de collaboration avec une entreprise tierce, cette clause sert à définir le périmètre d'action de chacun.

03 Une présentation des documents à numériser

Description détaillée des documents à numériser. Doivent notamment figurer dans cette section la nature des documents à numériser, leur histoire matérielle, leur état de conservation, leur support, leur taille et la quantité de documents. Les spécificités des documents doivent également être signalées : formats hétérogènes, fragilité des documents, reliures éventuelles, etc. On précisera également dans cette section si la numérisation doit être faite par double page ou simple page, et si toutes les pages du document doivent être numérisées.

04 La décomposition des documents à numériser sous forme de lots.

Selon l'hétérogénéité et la quantité de documents à numériser, il est indispensable de rassembler les documents en sous-ensembles. La taille et le nombre de ces lots doivent tenir compte de la capacité à les contrôler dans un délai raisonnable. La composition de l'échantillon de pré-production doit être définie lors de cette étape : ce lot servant de repère durant toute la phase de production, il doit refléter la variété de l'ensemble des documents numérisés.

05 Le déroulé de la prestation et les délais d'exécution

Il s'agit ici de définir la chaîne opératoire du projet de numérisation, le délai de réalisation de chaque tâche et des contrôles qualité associés à ces tâches. Doit figurer dans cette partie le déroulé en détail de la phase de pré-production. Si cette dernière n'est pas validée, la production en elle-même ne peut être démarrée. Ces tests doivent être réalisés dans les conditions réelles de production.

06 La cartographie de la chaîne de numérisation

Dans le même esprit que le point précédent, une représentation graphique de la chaîne de numérisation offre une vue d'ensemble intéressante à ajouter au cahier des charges : elle permet à chacun de mieux comprendre son rôle au sein du processus, d'appréhender l'enchaînement et les interactions entre les

étapes et de visualiser les étapes clefs dans la gestion de la qualité du projet.

07 Les conditions générales d'exécution

Préciser les règles de base qui devront être suivies tout au long du processus : les normes de numérisation suivies par l'institution, quelques grands principes sur la qualité des résultats attendus (notamment en termes de lisibilité ou d'esthétisme des contenus numérisés).

08 La description précise de la préparation des documents

Description de la manière dont les documents doivent être préparés (identification et marquage des documents, classement par formats, conditionnement, etc.).

09 La définition des modalités de transport et de stockage

L'importance de cette clause dépend évidemment de l'externalisation éventuelle de l'ensemble ou d'une partie du projet de numérisation. En cas d'externalisation, il convient de définir ici les modalités de prises en charge et de transport des originaux. Les contraintes de conservation et de stockage des documents doivent être clairement énoncées. Si l'ensemble de la chaîne de production se déroule localement, cette section permet de déterminer les modalités de stockage provisoire, ainsi que les conditions de manipulation des documents originaux.

10 L'étalonnage de la chaîne de numérisation

Décrire le type de procédure devant être mise en œuvre par l'atelier de numérisation pour paramétrer sa chaîne de numérisation avant le début de la production – en accord avec l'organisme commanditaire – et pour contrôler le maintien du réglage initial tout au long de la prestation.

L'étalonnage de la chaîne de numérisation porte sur tous les éléments qui peuvent faire l'objet de réglages automatiques : résolution, couleur, post-traitement, format de fichier. Il est effectué, dans les conditions normales de production, sur un ensemble de documents représentatif de la variété de ceux qui sont à traiter, puis analysé au regard des indicateurs demandés dans le cahier des charges, jusqu'à être validé conjointement. En même temps qu'il permettra de régler l'outil de numérisation, l'étalonnage servira également de référence tout au long de l'opération pour assurer une constante de qualité.

11 La démarche à suivre en cas de constat d'anomalie

Il peut s'agir de retirer le document du circuit de production, de signaler le cas pour analyse et résolution, de définir la nature du traitement approprié ou encore d'alimenter le système documentaire pour systématiser ces procédures.

12 Les caractéristiques des copies numériques à fournir

Définir les caractéristiques techniques (taille, couleur, format de fichier, cadrage, rapport d'agrandissement, résolution, colorimétrie, type et taux de compression, orientation, etc.) et les métadonnées techniques des images à fournir.

13 Le plan de nommage de tous les fichiers

Le plan de nommage correspond à l'ensemble de règles communes pour la dénomination des fichiers et doit être décrit précisément dans cette section.

14 La structure du fichier de récolement des métadonnées

Énumérer toutes les catégories devant figurer dans le fichier des métadonnées descriptives.

15 Les travaux d'indexation à réaliser

Définition des travaux d'indexation à réaliser et notamment les types d'informations à relever. Sera précisé dans cette clause le standard de métadonnées à suivre. La forme concrète de la saisie dépendra des modes d'accès et de recherche prévus : bases de données relationnelles, fichiers XML, etc.

16 Les modalités de contrôle et de suivi du fichier de récolement

Définir les informations qui doivent être fournies par le prestataire pour le suivi de la production : délai, forme sous laquelle le document doit être complété et remis, etc.

17 Les critères de contrôle des livrables

Cette section décrit l'étape finale du contrôle qualité⁴ en fin de chaîne, et correspond au contrôle d'intégrité de l'ensemble des produits de la prestation. C'est ici que l'on trouvera précisément le niveau de qualité attendue de la prestation et la marge de manœuvre dont disposera le commanditaire pour rejeter des produits non conformes à ses attentes. Seront également présents dans cette clause le nombre de contrôles effectués durant le processus de numérisation, de même que les délais pour contrôler les résultats intermédiaires et finaux.

18 La livraison des images numériques

En cas d'externalisation, cette clause permet de définir le mode de livraison des copies numériques, la restitution des originaux, la vérification des éventuels supports amovibles sur lesquels sont stockées les copies numériques (clé USB, disque dur externe, etc.).

⁴ Pour rappel, ce guide distingue deux types de contrôle qualité : le contrôle de la qualité comme démarche globale pour améliorer et contrôler l'ensemble du niveau de qualité de la chaîne de numérisation et le contrôle qualité, comme étape à part entière de la numérisation durant laquelle est vérifiée l'adéquation entre les objectifs initiaux du projet et le produit final obtenu.

POUR ALLER PLUS LOIN

Pour une vue exhaustive des lois, arrêtés royaux et circulaires régissant la question du **droit d'auteur en Belgique**, voir par exemple IPNews.be, site d'actualités juridiques spécialisé dans les nouvelles technologies.

<https://www.ipnews.be/legislation/legislation-belge/>

Le **Bureau de Coopération Interuniversitaire (BCI) du Québec** propose dans son **Guide de gestion d'un projet numérique** des outils d'évaluation, de planification et de contrôle utiles à la bonne conduite et à la réussite d'un projet de numérisation. Bien qu'initialement prévu pour des projets de numérisation de documents administratifs, l'angle d'approche de la numérisation se prête tout aussi bien à des archives historiques ou patrimoniales.

<https://www.enssib.fr/bibliotheque-numerique/documents/64628-guide-de-gestion-d-un-projet-de-numerisation.pdf>

Pour des exemples concrets de cahiers des charges, se référer au **Guide technique de la Direction des Archives de France - Écrire un cahier des charges de numérisation du Patrimoine**.

https://francearchives.fr/file/bf50d8fa5f554586dbf18fdc862d25970a1da0a7/static_4132.pdf

03.

LA QUALITÉ DES LIVRABLES : LES IMAGES

3.1 Image : définition et caractéristiques

Une image numérique désigne toute image acquise, traitée et stockée sous forme binaire qui se caractérise par les éléments suivants : la définition, la résolution, la profondeur de codage, le mode, l'espace colorimétrique et enfin le format.

LA RÉOLUTION ET LA DÉFINITION

La résolution, bien souvent associée à la qualité de l'image, car exprimant **la finesse d'analyse de la numérisation**, ne doit pas être confondue avec la définition de l'image. Cette dernière correspond au **nombre de pixels constituant l'image et se calcule à partir de la résolution et de la taille du document**. La définition est donc une valeur absolue permettant de quantifier une image et le capteur qui a servi à sa production. À noter que cette valeur seule ne permet pas de connaître les dimensions de l'image : un pixel n'a en effet

pas de taille « physique » en soi. Toutefois, dès lors que l'image est affichée ou imprimée, il faut prendre en compte la surface d'affichage, donc la résolution, qui est l'unité de mesure exprimant la finesse d'analyse de la numérisation. Elle s'exprime en ppp (pixel par pouce ou *pixel per inch*) ou dpi (*dot per inch*).

Il est communément admis qu'une **résolution minimale de 300 ppp à 100 % de l'original est adéquate pour la majorité des documents numérisés**, sauf exception pour les très grands ou petits formats. Ce standard tend à évoluer vers une résolution à 400 ppp pour un plus grand confort de lecture sur écran.

LA PROFONDEUR DE CODAGE

La profondeur de codage représente **la quantité de bits d'information qui définit un pixel**. Les bits d'information définissent le nombre de valeurs différentes qu'un pixel peut représenter.

Avec 8 bits, on atteint 256 valeurs différentes. Chaque bit additionnel double le nombre de valeurs possibles.

Pour une image en niveaux de gris, une profondeur de codage de 8 bits est adéquate. Pour une image en couleurs, une profondeur de codage de 24 bits est recommandée.

Pour une image en couleurs, certains appareils numérisent à une profondeur de codage supérieure, soit à 36 bits ou à 48 bits (12 ou 16 bits par couche).

LE MODE

Un mode est un modèle mathématique destiné à définir les relations des couleurs entre elles par le nombre de couches présent au sein de ces modes. Ces couches déterminent le niveau de détail perceptible de la couleur ainsi que la taille du fichier.

Le tableau ci-dessous¹ présente les trois modes principaux que l'on utilisera différemment selon le type de document à numériser :

NOM DU MODE	NOMBRE DE COUCHE	DÉTAIL DES COUCHE	APPLICATION
CMJN / CMYK	4	Cyan - Magenta - Jaune - Noir Attribue à chaque pixel un pourcentage pour chacune des couches. Bien que le modèle CMJN soit un modèle de couleur standard, la gamme exacte des couleurs représentées peut varier selon la presse et les conditions d'impression.	Mode privilégié pour l'impression offset. (Son utilisation n'est pas recommandée pour la numérisation.)
RVB / RGB	3	Rouge - Vert - Bleu Les valeurs de chaque pixel sont comprises entre 0 et 255 et permettent de reproduire plus de 16,7 millions de couleurs différentes.	Mode privilégié pour la numérisation en couleur.
Niveau de gris / Gray-scale	1	La valeur de chaque pixel est comprise entre 0 et 255, ce qui permet de reproduire 256 valeurs de luminosités allant du blanc pur (255) au noir pur (0).	Mode privilégié pour la numérisation en noir et blanc.
Certains modes ne sont pas envisageables dans le cadre des projets de numérisation comme les modes Couleurs indexées, Bitmap ou encore Bitonal.			

tab. 1 Les modes adaptés pour la numérisation patrimoniale

¹ Informations provenant du guide de numérisation de la BanQ et du site d'Adobe : Marie-chantal Anctil, Michel Legendre, Tristan Müller, Kathleen Brosseau, Louise Renaud, Recueil de règles de numérisation, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014.
<http://collections.banq.qc.ca/bitstream/52327/2426216/1/4671601.pdf>
<https://helpx.adobe.com/fr/photoshop/using/color-modes.html>

L'ESPACE COLORIMÉTRIQUE

Un espace colorimétrique est un **modèle mathématique tridimensionnel représentant l'ensemble des couleurs perceptibles, utilisables ou reproductibles au sein d'un mode**. Chaque couleur qu'il contient est ainsi associée à des coordonnées déterminant un point précis et correspondant à des valeurs telles que la luminance, la saturation et la teinte.

Il existe deux types d'espaces colorimétriques :

- **Les espaces dépendants** d'un appareil en particulier, également désignés sous le nom de **profil ICC**
- **Les espaces indépendants** décrivant un ensemble de couleurs visibles sans faire référence à un périphérique, également désignés sous le nom de **profil LCC**.

Quels que soient leur type, les espaces colorimétriques se différencient par leurs gamuts, c'est-à-dire par le nombre de couleurs accessibles et reproductibles au sein de cet espace. Parmi les espaces colorimétriques RVB les répandus, on trouve, du plus petit au plus grand gamut : sRVB, Adobe RVB (1998), ou encore le Spectre Visible, comme le montre la figure ci-contre.

À noter que les coordonnées spatiales varient d'un espace à l'autre et qu'un même ensemble de coordonnées correspond à plusieurs couleurs selon l'espace colorimétrique. Par exemple, sur l'illustration ci-dessous, le vert est associé aux coordonnées RGB [0 ; 255 ; 0] correspond à trois verts différents selon que ces coordonnées s'appliquent au spectre visible, à Adobe RVB ou à sRVB.

Dans le cadre de la numérisation, il est communément admis d'utiliser Adobe RVB 1998 pour une image couleur et Gray Gamma 2 pour une image en niveaux de gris².

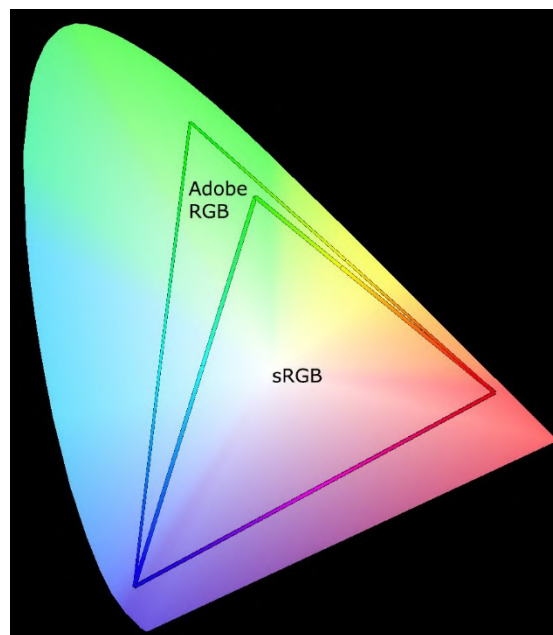


fig 2. Espaces colorimétriques RGB et sRGB

LE FORMAT

Le premier fichier image issu de l'étape de scannage est dit « RAW ». Cette désignation générique signifie que le fichier image obtenu contient toutes les données brutes du capteur, ainsi que les paramètres nécessaires à sa transformation en fichier image visible sur écran. Pour qu'il soit utilisable, il est nécessaire de **dématricer le fichier, c'est-à-dire de le convertir en un autre format**.

2 Marie-chantal Ancitil, Michel Legendre, Tristan Müller, Kathleen Brosseau, Louise Renaud, Recueil de règles de numérisation, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014.
<http://collections.banq.qc.ca/bitstream/52327/2426216/1/4671601.pdf>

Le format de sauvegarde du document numérisé varie en fonction des finalités du projet de numérisation (réalisé à des fins de diffusion ou de conservation).

Voir le tableau ci-dessous pour un aperçu des formats les plus pertinents selon le type d'exploitation envisagé :

TYPE DE DOCUMENT	FORMAT DE DIFFUSION	FORMAT DE CONSERVATION
Texte	.docx, .odt, .pdf (PDF) .epub .tei	.pdf (PDF/A)
Image	.jpg (JPEG), .jpg (JPEG2000), png	.tiff
Carte / Plans	.dwg, .svg	.dxf .svg

tab. 2 Formats à utiliser selon les exploitations possibles

Source : Sébastien Soyeux, *La numérisation en marche : les étapes de la dématérialisation des processus de travail*, Bruxelles : Archives générales du Royaume et Archives de l'État dans les Provinces, n.d.

http://www.arch.be/docs/brochures/la_numerisation_en_marche.pdf

Pour les fichiers images, le format **TIFF** est souvent privilégié, car d'excellente qualité. Ce format est cependant très lourd, et ne conviendra pas pour une consultation régulière.

Il est donc préférable disposer d'une première version TIFF qu'il sera ensuite possible de convertir dans un format approprié à la demande de l'utilisateur ou pour une exploitation interne. Mieux vaut alors privilégier les formats JPEG, JPEG 2000 ou PNG, dont la différence se situe au niveau de la compression

et donc du poids du fichier. Le JPEG opère une perte relative de qualité lors de la compression de l'image pour que celle-ci soit facilement diffusable, tandis que le PNG compresse l'image sans altération et sans perte de qualité.

Pour les fichiers textes, le format **PDF est le standard le plus communément admis**, car reconnu par l'ISO (Organisation Internationale de Normalisation). Il est utilisé aussi bien dans un objectif de conservation que de diffusion, car outre sa grande interopérabilité, il permet de rechercher et/ou sélectionner du texte dans un document.

L'océrisation (ou OCR pour *Optical Character Recognition* - Reconnaissance Optique de Caractères) est justement l'étape consistant à convertir une image en un document texte. Le document sera ensuite structuré selon un standard donné, par exemple ePub ou TEI (cf. encart à la page suivant).

À noter qu'**aucun format n'est conçu pour être permanent**. Pour garantir la pérennité des formats utilisés, l'université d'Harvard, en collaboration avec Jstor, a développé le logiciel open-source JHOVE³ qui permet d'identifier, d'analyser, et de caractériser les objets numériques.

Le logiciel est ainsi en mesure d'établir les propriétés du fichier et leurs adéquations éventuelles avec certains standards et normes sélectionnés par l'utilisateur. À partir de là, il est possible d'analyser la pertinence du document numérique, de corriger ses éventuelles anomalies ou d'envisager sa conversion en un autre format en cas d'obsolescence.

3 Open Preservation Foundation, *Démarrer avec JHOVE*, 2015. <https://jhove.openpreservation.org/getting-started/>

3.2 Critères d'une image de qualité

La réalisation d'images fidèles aux documents originaux peut, au moment de la capture, être un processus complexe où de nombreuses variables entrent en jeu, comme l'expertise et le jugement des équipes de numérisation. Pour épauler les institutions sur les aspects techniques du processus et fournir un cadre méthodologique à la fois rigoureux et réaliste, il existe plusieurs normes et documents regroupant les meilleures pratiques pour aborder toute la complexité de la qualité des images.

Parmi ces référentiels figurent le **guide FADGI**⁴ et le **guide Metamorphoze**⁵ sur lesquels sont basés les critères d'une image dite de bonne qualité, présentés ci-dessous et comprenant : la complétude, l'exposition, le contraste, la couleur, la mise au point, la résolution, le parallélisme, le flare, le vignettage, le bruit et enfin les artéfacts.

LA COMPLÉTUDE

Le premier critère de qualité d'un lot d'images est sa **complétude**. Complétude tout d'abord de l'ensemble – toutes les pages ont-elles été numérisées ? – mais aussi complétude des éléments de chaque image – la page a-t-elle été numérisée dans son entièreté ? Dans ce dernier cas, le contrôleur doit tenir compte des impératifs liés au recadrage. En effet, la seule manière de démontrer au lecteur que

4 Thomas Rieger, *Federal Agencies Digital Guidelines Initiative. Technical Guidelines for Digitizing Cultural Heritage Material. Creation of Raster Image File*, 2016, 99 p.
http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

5 Hans van Dormolen, *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012, version 1.0, 44 p.
http://www.imagingetc.com/images/Resources/Images/PDFs/DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

L'OCÉRISATION

OU LA RECONNAISSANCE OPTIQUE DE CARACTÈRES

L'océrisation (ou OCR pour Optical Character Recognition – Reconnaissance Optique de Caractères) est une technique de traitement de l'image visant à en extraire le texte pour faciliter la lecture de ce dernier et permettre à l'utilisateur d'effectuer aisément des recherches.

Le principe de l'océrisation se divise en plusieurs étapes qui comprennent :

- *Le redressement de la page,*
- *La binarisation du document, c'est-à-dire sa conversion en noir et blanc,*
- *La segmentation de l'image en zones de texte, puis en lignes, mots et caractères,*
- *Et enfin la classification de ces zones par la reconnaissance des caractères puis des mots.*

Les logiciels d'océrisation les plus courants sont basés sur le principe de deep-learning, c'est-à-dire sur l'amélioration des performances du logiciel par l'apprentissage de caractères et mots déjà rencontrés. Si une lettre ou un mot est inconnu pour le logiciel, la personne chargée de l'océrisation indique sa signification qui est ensuite mémorisée puis consignée automatiquement dans le dictionnaire de l'outil, qui devient ainsi plus autonome et performant.

Les difficultés de l'océrisation sont de plusieurs ordres. La qualité l'image numérisée d'abord, qui doit être suffisamment redressée et contrastée. L'état du document original joue également un rôle prépondérant dans l'océrisation : les défauts d'impression, la dégradation de l'encre ou encore les éventuelles lacunes rendent difficile la reconnaissance de caractères pour le logiciel comme pour l'opérateur et entravent le processus.

La structure peut également être problématique lorsqu'elle est trop complexe – plusieurs colonnes – que la page comporte des illustrations, des polices de très petite taille ou des ajouts d'écriture manuscrite. Enfin, les logiciels sont surtout orientés autour de la reconnaissance de l'alphabet latin, rendant encore difficile la transcription, par exemple, de l'alphabet arabe où la dissociation entre les lettres et les mots n'est pas aussi marquée. Pour autant, la recherche constante autour du procédé d'océrisation permet d'envisager les perspectives suivantes :

- Une meilleure automatisation des logiciels et une meilleure gestion des documents complexes – par le développement de méthodes, d'outils et de modèles mathématiques – affinent la segmentation de l'image en zones de texte ainsi que la gestion des documents comprenant à la fois du texte et des images⁶.

- L'amélioration de la reconnaissance d'écriture manuscrite. À titre indicatif, les projets de recherche HIMANIS⁷ et CLAMM⁸ – menés par l'Institut de Recherche et d'Histoire des Textes (IRHT-CNRS) – sont consacrés à l'analyse d'images et à la reconnaissance d'écriture manuscrite pour les médiévistes. Le développement de nouveaux logiciels de reconnaissance, de classification et de datation d'écriture manuscrite a déjà permis de lire et d'indexer des corpus comprenant plusieurs milliers de pages.

6 Tan Lu, Homogeneity models for image processing in the cultural heritage sector, [Thèse de Doctorat non publiée], Vrije Universiteit Brussel, 2020, 245 p.

7 IRHT - CNRS, Historical MANuscript Indexing for user-controlled (HIMANIS) <https://www.irht.cnrs.fr/fr/ressources/carnet-de-recherche/himanis>

8 IRHT - CNRS, Classification of Latin Medieval Manuscripts (CLAMM) <https://www.irht.cnrs.fr/fr/ressources/sites-web-outils-corpus/clamm>

l'ensemble de l'objet a été numérisé est de laisser autour de la page un bord noir, appelé **cadre patrimonial**. Ce cadre doit être suffisamment large pour un éventuel recadrage – les outils de redressement d'image rognent en effet plus ou moins 3% du nombre de pixels pour 1 degré d'inclinaison⁹.

Lorsqu'une page est manquante, la seule correction possible est l'insertion de cette page au lot existant par le biais d'une nouvelle numérisation.

La personne chargée du contrôle doit aussi être attentive à la **séquence des images**, c'est-à-dire à leur ordre d'apparition à l'écran. La séquence doit refléter ce que le lecteur découvrirait en consultant l'objet en salle de lecture. Les éventuelles erreurs de séquençage seront corrigées en adaptant la numérotation des pages. Il arrive que des feuilles volantes, des signets ou des annotations accompagnent un objet.

La personne chargée du contrôle doit alors vérifier avec le conservateur la pertinence de l'inclusion de ces objets au sein de la séquence et déterminer la place qu'ils doivent y occuper. S'agit-il d'une note de l'auteur, d'une page accidentellement désolidarisée de sa reliure ou d'une note bibliographique ? Selon le cas, les images devront être insérées ou annexées et faire l'objet d'une description à part entière.

9 Cette proportion, donnée à titre d'exemple, est exponentielle jusqu'à 45° et diminue ensuite de manière inverse jusqu'à 0% à 90°. Les directives Metamorfoze imposent, pour un objet non relié, de ne pas dépasser 2° d'inclinaison sur le dessous de l'image. Pour un objet relié, il est parfois impossible d'assurer une telle rectitude dans l'image. Il vaut mieux alors dès lors élargir encore l'espace autour de l'objet.

Source : Hans van Dormolen, Metamorfoze Preservation Imaging Guidelines. Image Quality, 2012, version 1.0, 44 p. http://www.imagingetc.com/images/Resources/Images/PDFs/DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

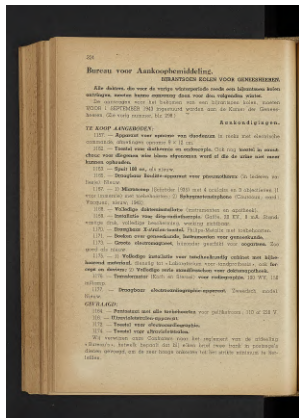


fig.3 Exemple de cadre patrimonial pour garantir la complétude du document numérisé.

Collection CegeSoma

L'EXPOSITION

L'exposition désigne l'action du rayonnement lumineux sur le capteur. Concrètement, une image surexposée est une reproduction numérique ayant capturé trop de lumière et qui apparaîtra donc trop claire, tandis qu'une image sous-exposée n'a capté que peu de lumière, et paraîtra trop sombre.

Il est très rare qu'une erreur d'exposition aille dans le sens d'une **sous-exposition**, qui se manifeste par une mauvaise lisibilité de l'image. La **surexposition** est par contre plus fréquente et plus grave, car elle peut entraîner une perte de données : les valeurs RVB d'un fichier numérique plafonnent à 255 ; si pour l'un ou plusieurs des canaux, la valeur atteint ce plafond, il devient impossible de corriger l'image, même *a posteriori*¹⁰.

Ajuster l'exposition d'une image peut se faire par le paramétrage de quatre éléments : la sensibilité du capteur, l'ouverture du diaphragme, la vitesse d'obturation ou la puissance lumineuse.

À noter qu'une trop grande sensibilité accroît le risque de bruit numérique. Avec l'expérience, la personne chargée du contrôle devrait être capable de déceler une surexposition de visu.

Il existe heureusement des outils pour l'aider dans cette tâche. Tout d'abord, l'**alerte de surexposition interne au scanner**, qui applique un masque de couleurs sur les zones impactées. Par défaut, cette alerte s'affiche dès qu'une des trois valeurs RVB atteint 250. Si aucune de ces valeurs n'atteint 255, il est encore possible de « récupérer » l'image, à condition de la dématriciser à nouveau – c'est-à-dire en créant un nouveau fichier TIFF sur la base du fichier RAW – en corrigeant l'exposition dans l'outil de capture.

Ensuite, un outil comme la « Pipette » proposée par le logiciel Photoshop peut par exemple aider à déterminer l'étendue du problème et à affiner l'analyse. S'il s'avère qu'une ou plusieurs images sont effectivement surexposées et qu'il y a donc une perte d'information, la personne chargée du contrôle devra rejeter les images et informer l'équipe de la nécessité d'une nouvelle capture.

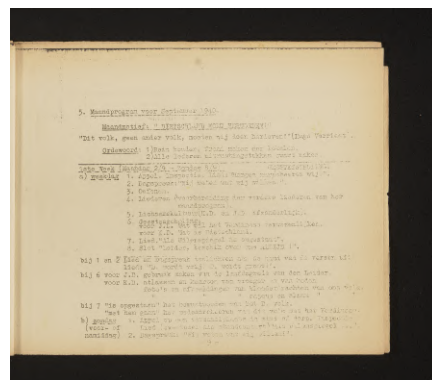


fig. 4 Exemple d'image sur-exposée

Collection CegeSoma

10 Cette situation souligne l'importance de la conservation des fichiers RAW, au moins jusqu'à l'archivage du fichier de conservation. En outre, il existe des systèmes de capture ne générant pas de RAW. Dans ce cas, toute erreur fera l'objet d'une nouvelle numérisation. Il est important d'en tenir compte lors de l'acquisition du matériel.

LE CONTRASTE

Le contraste d'une image correspond à **l'écart de densité entre ses teintes les plus claires et ses teintes les plus sombres**. Cette notion ne correspond pas à une distinction de type « clair ou foncé », mais bien au rapport entre ces deux types de teintes.

Le **contraste** joue un rôle important dans la lisibilité de l'image et doit dès lors être réglé avec soin. La **modulation de gain** est à cet égard un outil intéressant et consiste à **vérifier** – à l'aide de patches spécifiques lié à la charte colorimétrique – **le niveau de contraste selon les niveaux de gris et ainsi éviter une éventuelle saturation des couleurs**. Cette méthode de vérification fait l'objet de la norme **ISO 19264**¹¹ qui détaille son fonctionnement ainsi que les patches et niveaux de tolérance à respecter.

La mesure de la modulation de gain dans les hautes lumières permet donc de **valider le juste niveau des contrastes de l'image numérique**. L'objectif est de s'assurer de la lisibilité, et ce sans perte d'information, des tons doux et des teintes de papier jauni, ainsi que des encres dont le ton est proche de celui du papier.

Un mauvais paramétrage des niveaux de contraste peut rendre une encre illisible, surtout lorsque, brunie avec le temps, sa teinte se rapproche de celle d'un papier jauni. En cas de doute, la personne chargée du contrôle doit, bien entendu, **vérifier que les documents physiques sont effectivement déchiffrables**. S'il s'avère que le rendu à l'image n'est pas similaire à l'objet physique, il est nécessaire de vérifier que la fonction de modulation de gain est correcte.

Pour certifier ces valeurs, il est important que **chaque image créée contienne une charte**

standardisée reprenant les valeurs de gris depuis le blanc jusqu'au noir – charte de type Munsell Linear Gray Scale ou Kodak Gray Scale. Les mesures peuvent dès lors être rapidement prises. Dans le cas où les valeurs mesurées ne sont pas bonnes, le lot d'images doit être rejeté.

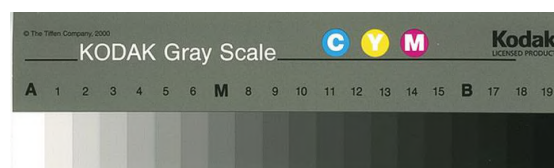


fig. 5 Charte KODAK Gray Scale



fig. 6 Charte Munsell Linear Gray Scale

La **fonction de modulation de gain** est pratiquement toujours représentée sous sa forme graphique et peut aussi abusivement être appelée gamma. On parle communément de **courbe de gradation**. Par défaut, sur les appareils reflex numériques, elle dessine une sigmoïde¹², notamment pour éclairer les parties les plus sombres et assombrir les hautes lumières afin de conserver les détails de toute la scène photographiée.

Dans le cadre d'une reproduction, dont le but est de proposer une version numérique fidèle à l'original, la courbe de gradation doit être linéaire. Contrairement à ce que l'on pourrait penser, ce réglage permet une meilleure lisibilité au niveau des contrastes doux, la

11 Organisation Internationale de Normalisation (ISO), ISO 19264-1:2021 - Photography – Archiving systems – Imaging systems quality analysis – Part 1: Reflective originals, juin 2021, <https://www.iso.org/obp/ui/fr/#iso:std:iso:19264:-1:ed-1:v1:en>

12 La courbe de gradation sigmoïde est un héritage de la photographie argentique et, en télévision, des caractéristiques du tube cathodique. En photographie, ce sont les qualités des films – sels d'argent, filtres, etc. – qui déterminent la courbe. Elle est intéressante à connaître pour le professionnel qui souhaite un résultat d'image particulier.

sigmoïde présentant une croissance très faible dans les blancs et les noirs profonds.

Il est possible d'adapter cette technique de mesure à toute la gamme des neutres jusqu'aux noirs les plus profonds à l'aide de la charte Munsell Linear Gray Scale dans le cadre d'une numérisation en noir et blanc – on parle alors d'analyse de la fonction de conversion opto-électronique ou OECF.

L'analyse permet aussi, en comparant les patches, de révéler un éventuel effet de flare : il va de soi qu'en choisissant délibérément de prioriser la qualité des contrastes dans les hautes lumières, les limites de la machine seront mises en évidence dans les basses lumières.

Si l'analyse de la charte Munsell Linear Grayscale révèle une non-linéarité de la courbe de gradation dans les tons plus sombres malgré une stricte application des recommandations énumérées ci-avant, seules les qualités du capteur peuvent être mises en cause¹³.

LA COULEUR

Le contrôle de la bonne reproduction des couleurs est probablement l'étape la plus délicate du contrôle de qualité tant sa perception peut varier d'un individu à l'autre¹⁴. En outre, la détection d'une erreur colorimétrique s'avère beaucoup plus difficile lorsque l'ensemble des images est faussé ou quand un document ne fait l'objet que d'une seule image. L'erreur est en effet plus simple à détecter s'il y a un changement brusque de teinte lors du passage en revue d'une série de pages.

La bonne gestion de la couleur dépend du choix des éclairants, du choix du scanner, du

choix de la charte colorimétrique, de l'étalonnage du matériel de scannage et enfin de la conversion du fichier RAW en un autre format. Dans ce cas, il est possible de corriger les fichiers erronés en repartant du fichier RAW – à condition que la qualité initiale de la calibration soit suffisante – et en le dématricant à nouveau, tout en vérifiant que les profils ICC et LCC et la balance des blancs des images soient identiques aux paramètres de l'image de référence.



fig. 7 Comparaison entre une mauvaise et une bonne reproduction des couleurs

Collection CegeSoma

LA MISE AU POINT

Malheureusement, **une image floue est toujours une image perdue** : certes, beaucoup de logiciels de traitement d'image permettent une correction de netteté, mais les algorithmes utilisés n'étant pas standardisés, le procédé ne garantit pas de rétablir la lisibilité parfaite du

13 Pour vérifier la qualité du capteur, ce dernier doit être mis à nu face à un éclairant capable – avec un IRC de 95% – de produire une série de flashes dont les puissances, doivent conserver une température de couleur identique et proche des 5000 Kelvins. Le test doit être répété neuf fois afin d'assurer la validité des résultats, et ce dans des conditions de température et d'humidité prédéfinies. L'analyse des résultats, documentés selon un schéma strict, montre l'étendue des densités reproductibles par le capteur et le respect de la linéarité de l'éclairant – voir ISO 7589, ISO 14524 et ISO 21550.

14 Il existe un certain nombre de troubles de la vision des couleurs dont la plus sévère, l'achromatopsie, est caractérisée par une absence totale de vision des couleurs. Entre une vision normale et une achromatopsie, les ophtalmologues recensent un éventail de troubles parfois très difficiles à diagnostiquer. Il pourrait donc être intéressant de tester la vue des opérateurs et des contrôleurs – le X-Rite Color Challenge, par exemple, est relativement simple et rapide.

fichier. D'aucuns considéreront qu'une image lisible, même légèrement floue, reste une image lisible. C'est sans compter sur l'interprétation électronique que le logiciel d'océrisation peinera à effectuer en raison de la perte d'information.

La netteté d'une image devrait pouvoir être validée par un processus électronique. Dans les faits, cela est inapplicable, étant donné la taille de certains documents au regard de la taille des capteurs, ainsi que de la configuration de certains berceaux à livre – en V, aspirants, etc. Il en va donc de l'œil de la personne effectuant le contrôle.

En cas de doute, il est important que ce dernier agrandisse l'image dans son format d'affichage maximum, c'est-à-dire à 100% ou à l'échelle 1 : 1 – ce qui signifie que chaque pixel de l'image est représenté par un pixel de l'écran. Bien souvent, **c'est en comparant une image considérée comme nette avec une image douteuse que la personne chargée du contrôle trouvera un repère.** Enfin, il y a lieu de **remettre en doute un système de capture lorsqu'il présente des irrégularités récurrentes quant à la netteté de l'image.** Ce problème peut être dû à deux facteurs : **le mouvement ou le dérèglement intempestif**, le flou de mouvement se produisant lorsque la vitesse d'obturation est trop élevée.

Dès lors, la moindre vibration – le simple fait que la machine se trouve par exemple dans un lieu de passage – peut avoir une influence et risque d'altérer la qualité de l'image. Le mouvement de miroir d'un reflex numérique de piètre qualité peut aussi être handicapant. **La solution serait alors de diminuer la vitesse d'obturation par l'augmentation de la puissance lumineuse – l'utilisation du flash électronique étant la meilleure option.**

Le second facteur pouvant être à la source de problèmes récurrents de flous est le **dérèglement d'un système**, c'est-à-dire le fait qu'entre le paramétrage et la prise de vue, un incident se produise. Par exemple, il arrive que certaines optiques, fatiguées, tolèrent difficilement la position verticale : le mécanisme n'étant plus stable, la mise au point change dans le temps et les images seront dès lors à refaire. Dans de tels cas, la personne effectuant le contrôle doit avertir le responsable du matériel afin que des **changements appropriés soient effectués.**

LA RÉOLUTION

Les problèmes de résolution sont le plus souvent de deux ordres : **le dérèglement de la résolution en cours de prise de vue ou une erreur de paramétrage avant même la première prise de vue.**

Dans le premier cas, la personne chargée du contrôle observera alors un **changement dans la taille de l'objet à l'image**, accompagné d'une **dégradation de la netteté.** La seule solution à apporter est de **numériser à nouveau les images mal calibrées, mais encore faut-il pouvoir déterminer quelles images sont corrompues.**

Pour ce faire, le contrôleur peut se servir de la **charte colorimétrique associée au matériel de scannage lors de la phase d'étalonnage.**

La charte Munsell comporte par exemple une règle graduée dont le centimètre doit correspondre à un nombre d'environ 118 pixels pour 300 PPI¹⁵. La qualité de la netteté est un élément supplémentaire permettant d'identifier le changement de résolution. À noter toutefois

¹⁵ À noter que cette mesure ne peut pas être effectuée avec toutes les chartes sur le marché : en effet, les chartes Kodak Q13/Q14, par exemple, contiennent bel et bien une règle, mais les mesures y sont inexactes. C'est le cas aussi pour les chartes Color Checker. Dans de tels cas, la mesure peut alors être effectuée avec l'outil « Règle » de Photoshop ou l'outil « Recadrage » de tout autre logiciel.

que le phénomène n'est pas toujours facilement détectable, car il est rarement la conséquence d'un changement brutal : il s'agit plutôt d'un élément du scanner – comme le banc de reproduction, le trépied, etc. – qui, mal fixé, va s'affaisser graduellement et progressivement détériorer la qualité des images.

Dans de tels cas, il est conseillé de **comparer la taille de la charte placée sur le fond de numérisation des premières et dernières images d'une série, afin d'identifier plus facilement un éventuel changement de dimension**. Ce point de contrôle est essentiel, car s'il n'est pas directement identifié à la fin de la phase de production, ce n'est que lors de la phase de recadrage que la personne chargée de la retouche des images s'en apercevra et toute la chaîne de numérisation s'en trouvera alors ralentie.

La seconde cause de problèmes de résolution, à savoir **une erreur initiale de paramétrage**, est plus difficile à détecter de visu, étant donné qu'il peut s'agir d'un mauvais ratio appliqué à toute une série d'images, voire à une image isolée. Un contrôleur expérimenté pourra, à condition de connaître parfaitement l'objet et le matériel, détecter ce problème à l'œil nu. Ici encore, c'est la règle graduée de la charte Munsell qui servira de point de contrôle.

LE PARALLÉLISME

La personne chargée du contrôle aura aussi pour tâche de vérifier que, lors de la prise de vue, le document soit parfaitement parallèle au capteur photographique. La chose est aisée lorsque l'objet numérisé présente au moins deux angles droits : il suffit d'appliquer à l'image des grilles de contrôle – présentes sur toutes les visionneuses – et de vérifier que les bords de la page adjacents à ces angles y sont alignés. Si des lignes de fuite

apparaissent à l'image, cela signifie qu'il y a divergence.

Si l'objet ne présente aucun angle droit, le contrôleur devra, en cas de doute, prendre au moins deux mesures sur des points précis et excentrés de l'objet avant de les comparer aux valeurs représentées à l'image sur les mêmes points. Une autre manière d'identifier une erreur de parallélisme est de vérifier si une zone de l'image est moins nette.

Pour ce faire, le contrôleur doit agrandir l'image à 100% et la faire défiler sur ses quatre côtés et en son centre. Cette méthode n'est toutefois pas toujours fiable, car certains capteurs de grand format induisent une telle profondeur de champ qu'une différence d'orientation de quelques millimètres restera imperceptible – d'autant plus que la sensibilité de l'œil varie d'un individu à l'autre.

En cas de problème de parallélisme, les images corrompues doivent impérativement être rejetées. Un technicien doit en être informé pour analyse et les images doivent être rescannées dans des conditions adéquates.

LE FLARE

L'effet de flare est une **aberration optique qui advient lorsqu'une lumière parasite pénètre dans l'objectif**. Ce dernier, ne devant théoriquement capter que les rayons reflétés par l'objet pour les guider au capteur, reçoit alors d'autres rayons, issus d'une source lambda.

Le phénomène est **caractérisé par un adoucissement du contraste sur l'ensemble de l'image, parfois par une distorsion chromatique, et, dans les cas les plus sévères, par un effet de halo**. Une fois détecté, l'enjeu est de trouver son origine afin d'annuler le phénomène. Lorsque le flare semble être perçu, le contrôleur doit comparer les patchs 0 et 5 de la charte Munsell

présente sur l'image. Si la différence de valeur RGB entre ces deux mesures est insignifiante, il doit en avertir le technicien, qui pourra déterminer l'origine du problème et la corriger avant de poursuivre la numérisation.

L'effet de flare est le plus souvent **provoqué par un mauvais positionnement de l'éclairant** – moins de 30° par rapport au plan de capture. Dans ce cas, il peut être aisément **compensé par l'utilisation d'un pare-soleil couvrant**. Si le phénomène persiste, il est intéressant de vérifier d'autres paramètres, tels que la **propreté de l'objectif** – la trace d'un doigt sur une optique suffit à provoquer le flare –, la **nature du sol de la pièce de numérisation** – un revêtement brillant par exemple – ou encore la **présence accidentelle d'une surface réfléchissante proche de l'objet**.

Les images présentant un effet de flare doivent être rejetées et le document doit à nouveau être numérisé.

LE VIGNETTAGE

Le vignettage est caractérisé par un **assombrissement des coins d'une image**. Il s'agit donc d'un **phénomène altérant l'uniformité de la lumière sur le plan de capture**.

Il est relativement difficile de le détecter, car l'objet couvre rarement l'entièreté du fond de numérisation. Parallèlement, ce fond étant noir pour améliorer la lisibilité de l'image et les coins de page d'un livre étant les plus exposés à la manipulation et donc à l'usure, ces zones peuvent déjà présenter un assombrissement perceptible sur l'objet physique, rendant plus difficile l'identification d'un problème de vignettage. Le contrôleur sera donc particulièrement attentif à l'état de l'objet en cas d'hésitation.

Le vignettage peut avoir différentes causes. La plus connue est une **incompatibilité entre l'optique et le capteur : l'optique, de par sa forme, envoie vers le capteur, rectangulaire, une image circulaire plus ou moins grande**. Cette image se nomme cercle image. Si le cercle image n'est pas capable de couvrir de manière uniforme l'ensemble du capteur, la luminosité sur les coins de l'image est diminuée. Cependant, dans le cadre de la numérisation patrimoniale, l'optique est censée avoir été testée et ne pas présenter ce défaut.

Il est plus probable, si un vignettage est détecté, qu'une lampe ait été déplacée par accident. Le contrôleur avertira dès lors le responsable du paramétrage des machines pour analyse. **En vérifiant les qualités de l'éclairant et en dématricant à nouveau quelques images, ce dernier sera en mesure d'identifier la source du problème.**

LE BRUIT

Le bruit numérique se définit comme la **présence d'informations sur l'image numérique qui n'apparaissent pas sur le document original**. Ces informations sont issues de la numérisation. Le rapport signal/bruit étant au plus fort dans les basses lumières, et le plus visible sur des zones de couleur uniformes, **c'est en observant l'arrière-plan noir d'une image que la personne chargée du contrôle pourra au mieux détecter le phénomène.**

Concrètement, il se traduit par **des ensembles de pixels dont les couleurs altèrent la lisibilité et la netteté de l'image**. Le plus souvent, le bruit est dû à la sensibilité trop élevée du capteur – ASA ou ISO – et désigné comme **bruit de lecture**. Le premier réflexe de la personne chargée du contrôle sera donc de **vérifier dans les métadonnées techniques de l'image si la valeur ISO est réglée au minimum.**

À titre indicatif, la majorité des capteurs sortant d'une chaîne de montage ont une sensibilité optimale entre 50 et 200 ISO. Les machines les plus récentes peuvent délivrer un rendu plus lisse encore jusqu'à 400 ISO¹⁶.

Aucun système ne permettant leur correction, **les images présentant du bruit numérique sont à rejeter et doivent à nouveau être numérisées.**

LES ARTÉFACTS

Enfin, il arrive par accident **qu'un artéfact, vienne perturber la bonne lecture de l'image ou simplement gâcher son esthétique.** Il peut s'agir de poussières sur une vitre, d'une charte de contrôle qui serait présente sur l'image recadrée ou d'un objet destiné à maintenir ouverte la page d'un livre.

Dans de tels cas, **tout dépend de la politique de numérisation mise en place par l'institution.** Ainsi, un outil visible permettant de redresser les lignes d'un texte aura tout son sens dans le cas d'une page imprimée destinée à l'océrisation, mais sera à exclure dans le cadre de la production d'un fac-similé d'un manuscrit.

Afin d'émettre une ligne directrice, il est toutefois possible d'établir que **tout ce qui peut être aisément retiré de l'image et n'est pas utile à sa lecture devrait être ôté.** Par exemple, une charte de contrôle qui empiète sur un objet risque de perturber l'océrisation – car elle contient des zones de texte.

Dans le même ordre d'idées, **les poussières ou résidus de papier présents sur les berceaux, les vitres ou le fond induiront, en fonction de la gravité de la situation et des objectifs d'exploitation, une nouvelle numérisation.**

La personne chargée du contrôle pourra attirer l'attention de l'opérateur sur le problème rencontré afin d'éviter qu'il ne se présente pas à nouveau. Si l'altération de l'image provient d'un défaut matériel, il en informera aussi l'opérateur technique, lequel prendra les mesures nécessaires pour faire à nouveau entrer la machine en production.

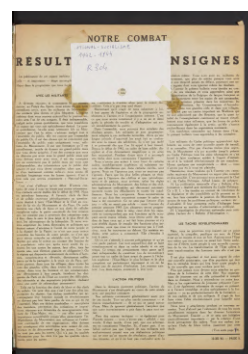


fig. 8 Exemple d'artéfact

Collection CegeSoma

16 Il est toutefois déconseillé de s'en approcher, car, au fil des prises de vue, les charges électriques répétées sur les photodiodes et le traitement informatique incessant peuvent induire un échauffement du système et altérer le signal de sortie, désigné cette fois comme un bruit thermique.

À noter qu'un capteur resté trop longtemps en mode live view va se mettre à chauffer. Il se peut donc qu'une série d'images soit affectée par le bruit thermique avant que le phénomène ne s'estompe avec la descente en température du capteur.

POUR ALLER PLUS LOIN

BIBLIOTHÈQUE NATIONALE DE FRANCE., *Techniques et formats de conversion en mode texte*, BnF - Site institutionnel., n.d, [En ligne],
<https://www.bnf.fr/fr/techniques-et-formats-de-conversion-en-mode-texte>

CHIRON, G., DOUCET, A., COUSTATY, M., VISANI, M., and MOREUX, J.-P.,
« Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information », *In: 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2017, pp. 1-4.

LU, T., *Homogeneity models for image processing in the cultural heritage sector*, [Thèse de Doctorat non publiée], Vrije Universiteit Brussel, 2020, 245 p.

MEMON, J., SAMI, M., KHAN R.-A., UDDIN M., « Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR) », *In : IEEE Access*, vol. 8, blz. 142642-142668, 2020.

RIEGER, T., *Federal Agencies Digital Guidelines Initiative. Technical Guidelines for Digitizing Cultural Heritage Material. Creation of Raster Image File*, 2016, 99 p.,
[En ligne],
[http://www.digitizationguidelines.gov/guidelines/
FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-
2016%20Final_rev1.pdf](http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf)

VAN DORMOLEN, H., *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012, version 1.0, [En ligne],
[http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/
Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf](http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf)

04.

LA QUALITÉ DES LIVRABLES : LES MÉTADONNÉES

4.1 Métadonnées : définition et caractéristiques

Les métadonnées, comprises au sens large, sont **des données sur des données**. Elles ont pour but de **définir ou de décrire une autre donnée, et ce quel que soit son support** – papier ou numérique.

Déjà utilisées avant l'ère numérique par les bibliothèques et les musées dans le cadre des catalogues et des inventaires papier, le terme « métadonnée » s'est vu généralisé avec l'informatisation croissante de nos sociétés. Aujourd'hui, le terme correspond à « **l'ensemble des informations décrivant une ressource quelconque, numérique ou non [et dont] la fonction première est de décrire le contenu de la ressource, tout en permettant de l'identifier, de le qualifier et de l'enrichir¹** ».

De la même manière qu'il existe un large éventail de contenus patrimoniaux et documentaires aux caractéristiques variées, il existe différents types de métadonnées. Ces données s'inscrivent de

La fonction première [des métadonnées] est de décrire le contenu de la ressource, tout en permettant de l'identifier, de le qualifier et de l'enrichir.

façon **interne** ou **externe** au fichier numérique. Les métadonnées sont dites internes lorsqu'elles sont **encapsulées au sein du même fichier** que la source numérisée. À l'inverse, les métadonnées dites externes sont **contenues dans un fichier ou une base de données séparée du document numérisé**.

¹ Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, 118 p.
https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf

Il est possible d'en distinguer sept catégories :

TYPE	DESCRIPTION	EXEMPLES
Identifiantes	Identifient formellement le contenu numérisé, le plus souvent sous forme de chaîne alphanumérique.	<ul style="list-style-type: none"> ● ISBN ● ISNI ● URI
Descriptives	Décrivent le plus précisément possible et de manière objective le document numérisé.	<ul style="list-style-type: none"> ● Nom de l'auteur ● Titre de l'œuvre ● Matériaux ● Date de création
Administratives	Renseignent sur la date de création du fichier et ses modifications ultérieures.	<ul style="list-style-type: none"> ● Date de création du fichier ● Nom de l'institution créatrice du fichier ● Identification du document source
Juridiques	Précisent toutes les modalités juridiques auxquelles sont assujetties les contenus numérisés : droit d'auteur, ayant-droit, licence d'utilisation.	<ul style="list-style-type: none"> ● Nom du titulaire ou du gestionnaire du copyright ● Mention de sources devant figurer avec le titre de l'œuvre
D'enrichissement	Fournissent des informations supplémentaires sur l'œuvre ou le document.	<ul style="list-style-type: none"> ● Biographie de l'artiste ● Photographies ● Notes ● Image de la couverture
Techniques (ou structurelles)	Renseignent sur le mode de création du fichier, son format et ses modalités d'utilisation.	<ul style="list-style-type: none"> ● Identifiant du logiciel et de sa version ● Mode de numérisation ● Format de compression ● Taille du fichier
D'usage	Informent sur la consommation d'un contenu culturel en particulier (le plus souvent constitués de manière automatique par la plateforme de diffusion du contenu numérisé).	<ul style="list-style-type: none"> ● Nombre de vues ● Recommandations faites par les usagers

tab. 3 Les différents types de métadonnées

Source : Observatoire de la culture et des communications du Québec, *État des lieux sur les métadonnées relatives aux contenus culturels*, Québec : Institut de la statistique du Québec, 2017, p. 19
https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf

4.2 Rôle et fonction des métadonnées

Les métadonnées servent à **organiser et à trier l'information**. Grâce à elles, il est possible d'interroger des catalogues, des répertoires et des bases de données pour trouver l'information recherchée. **La production de métadonnées n'est donc ni une stratégie ni une fin en soi**, mais bien un moyen pour répondre aux besoins et attentes des différents intervenants de la chaîne de valeur des produits culturels.

Plus l'information est structurée, plus il sera facile de la traiter et de la rendre pertinente. Face au flux de contenus, de publications et plus largement d'informations toujours plus important, les métadonnées garantissent également que l'information ne soit tout simplement pas perdue.

Avant de produire des données, il est donc essentiel de **connaître le contexte dans lequel l'information sera publiée et l'objectif de cette publication**. Il devient ainsi possible de déployer des liens qui relient notre information à toute autre information pertinente. Plus il y a de liens entre des œuvres, des personnes, des organisations, des lieux et des événements, plus le potentiel est grand, pour toutes ces informations, d'être trouvées et (ré)utilisées.

Pour répondre à ces besoins de manière pertinente, les métadonnées doivent satisfaire à certains prérequis : elles doivent ainsi être **interopérables, satisfaire à un niveau de qualité acceptable, et enfin découler d'une vision partagée par les différents intervenants et utilisateurs impliqués**.

4.3 Prérequis à l'utilisation des métadonnées : l'interopérabilité

La nécessité de structurer l'information est apparue dès les années 1970, avec la généralisation de l'informatique dans nos sociétés². Un des éléments-clés dans l'évolution des métadonnées a été la nécessité de disposer d'**informations lisibles et exploitables par les différents protagonistes d'un même secteur, qu'il s'agisse d'êtres humains ou de machines**.

C'est tout l'enjeu de l'**interopérabilité**, définie comme « la capacité que possède un produit ou un système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs, et ce sans restriction d'accès ou de mise en œuvre³ ».

Avant d'aller plus loin, il est important de **différencier l'interopérabilité de la compatibilité** : la compatibilité permet à deux systèmes de communiquer ensemble tandis que l'interopérabilité permet à deux systèmes de fonctionner ensemble.

La compatibilité porte entre un format A et un format B, et consiste à ce que deux éléments se comprennent : le format A des données d'un logiciel permet de recevoir l'information contenue dans le format B des données d'un autre logiciel et inversement. En somme, la compatibilité résulte d'un travail de traduction d'un format vers un autre, au cas par cas, notamment lorsque les formats sont fermés.

2 Thierry Claerr et Isabelle Westeel, *Numériser et mettre en ligne*. Villeurbanne : Presses de l'enssib, 2010. <http://books.openedition.org/pressesenssib/414>

3 Définition de l'interopérabilité, n.d., <http://definition-interoperabilite.info/>

L'interopérabilité en revanche est générale, puisqu'elle **s'appuie sur un ou plusieurs formats ouverts** (par exemple : l'HTML) **indépendants du logiciel de création et d'utilisation des données**. Ces normes ont un double rôle : elles indiquent la façon dont le dialogue entre les différents éléments va s'opérer et elles permettent une passerelle de communication qui sera en mesure de s'adapter aux besoins changeants des éléments. Plus spécifiquement, l'interopérabilité concerne trois aspects de l'échange d'informations :

L'INTEROPÉRABILITÉ TECHNIQUE

L'interopérabilité technique ou informatique concerne les modalités de liaison entre systèmes, la définition des interfaces, le format des données et les protocoles. Elle **décrit la capacité pour des technologies différentes à communiquer et à échanger des données basées sur des normes d'interface clairement définies en amont et univoques** comme par exemple HTTP ou OAI-PMH.

L'INTEROPÉRABILITÉ SÉMANTIQUE

L'interopérabilité sémantique **assure quant à elle que la signification exacte des renseignements échangés soit comprise par n'importe quelle autre application**, même si celle-ci n'a pas été conçue initialement dans ce but précis. Pour cela, les systèmes doivent partager un modèle commun d'échange d'informations, tels que les modèles de métadonnées Dublin Core ou MODS.

L'INTEROPÉRABILITÉ SYNTAXIQUE

Enfin, l'interopérabilité syntaxique concerne la façon dont sont codées et formatées les données, en définissant la nature, le type ou encore le format des messages échangés. Les langages XML et RDF sont des exemples de standards garantissant l'interopérabilité syntaxique.

INTEROPÉRABILITÉ DES IMAGES

LE PROJET IIIF

*Si de nombreux schémas et protocoles ont vu le jour afin de répondre aux besoins d'échanges de métadonnées, **les images sont longtemps restées en dehors de ces considérations**. En effet, la mise en ligne d'image de qualité reste encore aujourd'hui un processus complexe, technique et coûteux. Si les institutions culturelles ont par exemple créés des outils de visualisation des contenus numériques, ces derniers ne sont que trop rarement interopérables, emprisonnant les images sur leur site d'origine et ne les rendant consultables que par ce biais, ce qui les rend donc difficilement partageables et réutilisables.*

*C'est pour répondre à cette problématique qu'est né le **projet IIIF** (International Image Interoperability Framework) dont l'objectif est de **définir un cadre d'interopérabilité pour les bibliothèques numériques**. Cet ensemble de standards permet alors aux institutions scientifiques et culturelles de promouvoir leurs contenus « de manière standardisée sur le Web et de les rendre ainsi consultables, manipulables et annotables par n'importe quelle application ou logiciel compatible⁴». Ce cadre d'interopérabilité s'appuie sur deux piliers :*

- *Un modèle de données, le **Shared Canvas**⁵, basé sur des données liées pour décrire des contenus numérisés de manière collaborative.*
- *Des **APIs** (Application Programmable Interface), c'est-à-dire des ensembles de fonctions permettant d'accéder aux services d'une application, par l'intermédiaire d'un langage de programmation.*

Ces APIs permettent par exemple de définir une syntaxe d'URL standardisée permettant de

manipuler une image à distance ou encore de délivrer de manière standardisée des informations de présentation et de structure d'un objet numérique.

L'ensemble de ces informations sont rassemblées au sein d'un seul fichier, le « **manifeste IIIF** » qui en plus de décrire l'image numérique, décrit précisément la structure interne du contenu numérisé. Au sein de ce manifeste, ces informations sont couplées à un canevas, c'est-à-dire à un espace vide représentant de manière abstraite une vue particulière d'un contenu numérisé – une page dans le cadre d'un livre par exemple. Ce canevas va servir de réceptacle à toutes sortes de ressources numériques conçues comme des annotations sur ce même canevas.

Ce manifeste doit ensuite être lu par un visualisateur d'image compatible avec IIIF⁶, qui sera alors **capable de charger à distance les informations contenues dans le manifeste pour recomposer le contenu numérique dans une autre interface web que celle d'origine, pour ensuite pouvoir le manipuler et l'annoter à volonté.**

Les technologies IIIF ouvrent donc de nouveaux horizons pour valoriser le patrimoine culturel numérisé dans un environnement décentralisé par nature. Pour les institutions culturelles et scientifiques, les avantages de telles technologies sont nombreux : économiquement le projet n'est pas écrasant puisque porté par une large communauté internationale. La dépendance à un outil ou logiciel spécifique est moindre et il devient plus facile de changer d'outil et de choisir le meilleur dans sa catégorie. Plus largement, les standards IIIF augmentent le potentiel des contenus des bibliothèques numériques.

4 Régis Robineau, Comprendre IIIF et l'interopérabilité des bibliothèques numériques | Insula, 8 novembre 2016. <https://insula.univ-lille3.fr/2016/11/comprendre-iiif-interoperabilite-bibliotheques-numeriques/>

5 Robert Sanderson, Benjamin Albritton, Shared Canvas Data Model, n.d. <https://iiif.io/model/shared-canvas/1.0/>

6 La communauté IIIF a rassemblé sur Github tous les visualisateurs d'image compatibles avec les technologies IIIF : <https://github.com/IIIF/awesome-iiif#image-viewers>

4.4 Critères d'une métadonnée de qualité

Comme pour les images, les professionnels de la numérisation et les scientifiques ont identifié les critères permettant de définir une métadonnée comme étant de qualité. La catégorisation présentée ci-dessous est directement inspirée des travaux de l'Europeana Task Force⁷, du groupe de recherche FORCE11 et de leur FAIR principes⁸, des travaux de l'Observatoire de la Culture et des Communications du Québec⁹, de la *Quality Assurance Framework (QAF) for statistical data*¹⁰ développée par Statistics Canada (STC), et enfin des travaux des scientifiques de Thomas R. Bruce et Diane I. Hillmann sur le sujet¹¹.

Pour être de bonne qualité, les métadonnées doivent être :

01 Complètes

Les métadonnées doivent décrire aussi précisément que possible l'objet numérique. Ces catégories doivent être pensées pour pouvoir s'appliquer tant que possible à l'ensemble des fichiers numériques de la collection.

- 7 Marie-Claire Dangerfield, Lisette Kalshoven, Report and Recommendations from the Europeana Task Force on Metadata Quality, mai 2015, 54 p. https://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf
- 8 Force 11, Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0 <https://www.force11.org/fairprinciples>
- 9 Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, 118 p. https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf
- 10 European Statistical System, Quality Assurance Framework, n.d., version 2.0, 58 p. <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>
- 11 Thomas Bruce, Diane Hillmann, The Continuum of Metadata Quality: Defining, Expressing, Exploiting, 2004. https://www.researchgate.net/publication/247818823_The_Continuum_of_Metadata_Quality_Defining_Expressing_Exploiting

02 Exactes

Les métadonnées doivent être correctes et basées sur des connaissances du document objectives et factuelles. Les problèmes d'exactitude peuvent se manifester par des erreurs typographiques, des utilisations non conformes de noms, de lieux ou par la multiplication des abréviations.

03 Fiables de par leur provenance

La fiabilité des métadonnées dépend partiellement de la personne qui les a encodées, de son degré de compétences, de sa méthodologie pour créer et traiter les métadonnées, et enfin de son expérience. Il arrive qu'une autre personne modifie les métadonnées après la création de l'objet.

Il existe néanmoins des outils et méthodologies pour garantir la fiabilité des métadonnées : dans le secteur des bibliothèques par exemple, les Américains ont par exemple développé l'Anglo-American Cataloging Rules (AACR2).

04 Conformes aux attentes

Afin d'être significatives, les métadonnées doivent refléter les attentes et besoins des utilisateurs de l'objet numérique décrit. Pour cela, il est essentiel de cerner précisément l'objectif d'exploitation du chantier de numérisation et d'identifier les besoins du public cible de ce chantier.

05 Cohérentes

Pour garantir la cohérence des documents numérisés, le vocabulaire utilisé doit être régi par des règles précises pour éviter les inexactitudes.

06 Accessibles

Les métadonnées doivent être lisibles à la fois par les machines et par les humains. L'usage d'un standard trop complexe ou difficilement lisible entravera l'encodage et le traitement du document décrit. A fortiori, cette difficulté pourrait mener à des erreurs de cohérence et d'exactitude.

L'accessibilité se caractérise également par la facilité à découvrir l'objet en ligne, et correspond de fait à la problématique de l'identifiant. Ce dernier se doit d'être unique et pérenne. À titre indicatif, le standard d'identification URI¹² (Uniform Resource Identifier) répond à cette double exigence.

12 Un URI est une courte chaîne de caractères identifiant une ressource sur un réseau (par exemple, une ressource Web), physique ou abstraite, et dont la syntaxe respecte une norme d'Internet mise en place pour le World Wide Web. Le sigle URI est généralement utilisé pour désigner une telle chaîne de caractères. Les URIs sont la technologie de base du World Wide Web car tous les hyperliens sont exprimés sous cette forme. Ils se divisent en deux catégories: les URN et les URL.

Pour une vue générale des identifiants du Web recommandé par le W3C, consultez : <https://www.w3.org/Addressing/>

4.5 Une vision commune pour une meilleure interopérabilité

Il n'existe pas un standard universel suffisamment exhaustif pour rencontrer l'ensemble des critères de qualité précédemment cités tout en s'adaptant parfaitement aux besoins des institutions culturelles et scientifiques et de leurs publics. En pratique, les schémas de métadonnées ont souvent de multiples fonctions, car chaque type d'institution culturelle – bibliothèques, archives et musées – ont des besoins qui leur sont propres en matière d'identification et de catalogage de leurs collections.

Les **bibliothèques** ont une longue tradition du catalogage de leurs collections. Chaque ouvrage doit être précisément identifié grâce à une notice bibliographique, mais ces métadonnées ne sont le plus souvent pas très détaillées ou contextualisées. Eulalia Roel souligne à cet égard que « la force du modèle de la bibliothèque [réside dans le fait] qu'une quantité minimale de métadonnées est attribuée à un pourcentage très élevé d'objets dans la collection¹³».

L'un des enjeux majeurs des bibliothèques a été de faciliter l'échange des données bibliographiques entre bibliothèques et services d'archives. Cette problématique a notamment trouvé une réponse avec le format MARC de la Library of Congress, dont l'architecture et le fonctionnement facilitent ces échanges. De ce modèle a découlé plusieurs variantes : la BnF a par exemple développé le format InterMarc, tandis que le Canada a créé le CAN/Marc. En réalité, ces efforts d'individualisation ont paradoxalement complexifié les échanges de données, entraînant l'obligation de développer des programmes de conversion entre les standards de métadonnées.

Les **services d'archives** diffèrent des bibliothèques en ce sens que les métadonnées doivent avant tout contextualiser les différents objets de la collection en identifiant précisément les relations d'une collection avec des personnes, des lieux, etc. La difficulté majeure d'une telle démarche découle selon Eulalia Roel de la « grande incohérence dans les détails et les variables de description, ainsi que des niveaux de granularité très disparates dans la hiérarchie de la structure d'une collection à laquelle les métadonnées sont attribuées¹⁴». De telles disparités contribuent à des malentendus entre institutions lorsqu'elles tentent d'utiliser les métadonnées dans plusieurs disciplines.

Les **musées** se démarquent par une utilisation plus réservées des normes descriptives pour leurs collections, notamment car la culture des musées ne favorise pas la mise à disposition gratuite au public de toutes les informations relatives à une collection. Parallèlement, l'éclectisme des collections muséales rend souvent l'exercice extrêmement complexe : la rigidité des métadonnées serait discriminatoire à l'égard des informations nuancées requises pour chacun des objets d'une collection¹⁵.

Si l'interopérabilité a depuis bien longtemps été identifiée comme l'élément-clé dans les échanges de données entre institutions et entre secteurs¹⁶, les moyens pour atteindre ce niveau d'interopérabilité restent incertains.

13 Eulalia Roel, *The MOSC project: Using the OAI-PMH to bridge metadata cultural differences across museums, archives, and libraries*. *Information Technology and Libraries*, 24 (1), 2005, 22-24.

14 *Idem*

15 *Idem*

16 Kalina Sotirova, Juliana Peneva, Stanislav Ivanov, Rositza Doneva, Milena Dobrova, *Digitisation of Cultural Heritage : Standards, Institutions, Initiatives*, In : *Access to Digital Cultural Heritage : Innovative Applications of Automated Metadata Generation*, Plovdiv : Plovdiv University Publishing House, 2012. <http://www.math.bas.bg/infres/book-ADCH/ADCH-ch1>

Comme l'ont souligné Milena Dobreva¹⁷ et Philip Hider¹⁸, cette difficulté tient au fait que **l'interopérabilité est abordée selon deux axes** qu'il n'est pas aisé de conjuguer : d'une part, certaines institutions se concentrent sur **l'interopérabilité interne**, tandis que d'autres privilégient le **développement inter-institutionnel**. D'autre part, en ce qui concerne la manière de parvenir à l'interopérabilité, certaines institutions mettent l'accent sur le **respect des normes en matière de métadonnées**, tandis que d'autres insistent sur **la manière dont les nouvelles technologies peuvent fonctionner avec des formats de métadonnées et des contenus divergents**¹⁹.

L'objectif est donc aujourd'hui de développer des modèles de métadonnées qui permettent de partager les contenus numérisés entre les trois secteurs. Cette mutualisation passe par plus de souplesse et d'inclusivité dans leur interopérabilité. Sans prétendre à un panorama exhaustif des initiatives développées aux cours des dernières années, voici quelques modèles qui vont dans ce sens :

Le **CIDOC Conceptual Reference Model**²⁰ (**CIDOC CRM**) est un modèle conceptuel pour la description des objets muséaux développé depuis 2006 par le Comité international pour la documentation (CIDOC) du Conseil international des musées (ICOM). Il a pour objectif de **mettre à disposition toute l'information nécessaire à la documentation et à la gestion du patrimoine culturel en fournissant un cadre sémantique commun et extensible**, avec lequel toute information sur le patrimoine culturel peut être mise en correspondance. Toujours en évolution, il comporte à ce jour 86 classes et 137 propriétés et est centré autour de la notion d'événements, c'est-à-dire du type de relation qui unit un objet et ses caractéristiques — comme sa date de création par exemple.

Développé en 2017 par la Fédération Internationale des Associations et Institutions des bibliothèques (IFLA), l'**IFLA LRM**²¹ propose quant à lui un modèle pour les bibliothèques orienté autour des besoins des utilisateurs. Ce modèle **favorise en effet la structuration des données pour la recherche d'informations bibliographiques sur le Web** et permettrait par exemple aux utilisateurs d'identifier en priorité des œuvres ou des lieux avant de localiser une ressource.

Pour répondre à cet objectif, le modèle s'appuie sur trois notions : **l'entité**, c'est-à-dire l'objet décrit, **l'attribut**, qui correspond aux différentes caractéristiques de cet objet, et **les relations**. Ces relations concernent aussi bien une entité avec ses propres attributs que les relations entre plusieurs entités d'un ou plusieurs systèmes.

L'objectif est donc de développer des modèles de métadonnées qui permettent de partager les contenus numérisés entre les [...] secteurs [culturels].

À terme, **l'objectif est de remplacer les notices bibliographiques par un réseau de relations entre des entités et ainsi de favoriser leur visibilité sur le Web et leur exploitation par des machines.**

17 *Idem*

18 Philip Hider, *Australian digital collections: Metadata standards and interoperability*, Australian Academic & Research Libraries, 35, 2004.

19 *Idem*

20 Home | CIDOC CRM, <http://www.cidoc-crm.org/>

21 IFLA, IFLA Library Reference Model (LRM), <https://www.ifla.org/publications/node/11412>

Records in Contexts²² (RiC) est un projet de refonte des normes de description archivistique initié en 2016 par le Conseil International des Archives (ICA). Son objectif est d'améliorer et d'augmenter l'interopérabilité des descriptions archivistes publiées sur le Web en créant un modèle réunissant les normes de description archivistes existantes, et notamment ISAD(G)²³, ISAAR (CPF)²⁴, ISDF²⁵ et ISDIAH²⁶.

Comme les deux modèles précédemment cités, RiC propose également la mise en place d'un système relationnel entre les entités et leurs caractéristiques, de même qu'entre ces entités.

L'intérêt de ces nouveaux modèles de métadonnées réside dans **l'approche qu'ils ont des objets numériques**, considérées non pas comme des éléments isolés auxquels il faut simplement assigner un lot de caractéristiques descriptives, mais bien comme **un ensemble hétéroclite d'objets qui n'a réellement de sens que par le biais des relations liant ces objets les uns aux autres**. En réunissant des normes déjà existantes, en se basant sur le langage RDF (*Resource Description Framework*) et enfin en s'appuyant sur l'utilisation de données ouvertes et liées – *Linked Open Data* ou LOD en anglais – ces nouveaux modèles s'inscrivent pleinement dans la transition vers le Web sémantique.

En effet, cette approche permet non seulement de **constituer un réseau d'objets numériques plus cohérent et significatif**, mais également d'apporter **plus de visibilité à ces objets**. Pour les institutions culturelles et scientifiques, c'est la perspective de création et de mise en commun de corpus riches et complexes ouvrant la voie à une valorisation accrue de leurs collections.

Toutefois, si ces modèles œuvrent à plus de liens et d'interopérabilité, la manière dont les différents secteurs vont pouvoir combiner leurs

Pour les institutions culturelles et scientifiques, c'est la perspective de création et de mise en commun de corpus riches et complexes ouvrant la voie à une valorisation accrue de leurs collections.

données et interagir entre eux reste encore à définir.

Des premiers efforts d'harmonisation ont été faits dans ce sens avec, par exemple, le **projet FRBROO**²⁷, qui cherche à aligner voire à fusionner les modèles CIDOC CRM, principalement pensé pour les musées, et FRBR, développé quant à lui pour les bibliothèques.

22 International Council on Archives, Records in Contexts, Conceptual Model, n.d.

<https://www.ica.org/en/egad-ric-conceptual-model>

23 Norme assurant la rédaction de descriptions archivistes compatibles, pertinentes et explicites, tout en permettant l'utilisation de données d'autorités communes pour rendre possible l'intégration de descriptions provenant de différents lieux de conservation.

Source : Observatoire de la culture et des communications du Québec, État des lieux sur les métadonnées relatives aux contenus culturels, Québec : Institut de la statistique du Québec, 2017, p.98.

https://bdso.gouv.qc.ca/docs-ken/multimedia/PB01600FR_MetadonneesCulturel2017H00F00.pdf

24 Norme Internationale sur les notices d'autorité utilisées pour les Archives relatives aux collectivités, aux personnes ou aux familles.

Source : International Council on Archives, ISAAR (CPF), 2011.

https://www.ica.org/sites/default/files/CBPS_Guidelines_ISAAR_Second-edition_FR.pdf

25 Norme internationale pour la description des fonctions et des collectivités associées à la production et à la gestion des archives.

Source : International Council on Archives, ISDF, 2011.

https://www.ica.org/sites/default/files/CBPS_2007_Guidelines_ISDF_First-edition_FR.pdf

26 Norme internationale pour la description des institutions de conservation des archives.

Source : International Council on Archives, ISDIAH, 2011.

https://www.ica.org/sites/default/files/CBPS_2008_Guidelines_ISDIAH_First-edition_FR.pdf

27 Introduction à FRBROO | Documentation, n.d.
<https://doc.bibliissima.fr/introduction-a-frbroo>

POUR ALLER PLUS LOIN

CARBONE, F., OOSTERBEEK, L., COSTA, F., « Extending and adapting the concept of quality management for museums and cultural heritage attractions: A comparative study of southern European cultural heritage managers' perceptions », *In: Tour Manage Perspect*, 2020, 10 p.

FLECKENSTEIN, M., FELLOWS, L., « Data Quality », *In : Modern Data Strategy*, Cham: Springer, 2018.

BACHI, V., FRESA, A., PIEROTTI, C., PRANDONI, C., « The digitization age: mass culture is quality culture. challenges for cultural heritage and society », *In: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (eds.) EuroMed 2014*.

GUEGUEN, G., MARQUES DA FONSECA, V.-M., D., PITTI, V. et SIBILLE, C., *Vers un modèle conceptuel international pour la description archivistique. Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description*, 2013, [En ligne],
https://www.ica.org/sites/default/files/EGAD_French.pdf

CHARDONNENS, A., *La gestion des données d'autorité archivistiques dans le cadre du Web de données*, [Thèse de doctorat non publiée], Université Libre de Bruxelles, 2021, 420 p.
<https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/315804/Holdings>

CARON, P., DEFIOLLE, R., et Lay, M.-H., *L'enjeu des métadonnées dans les corpus textuels : Un défi pour les sciences humaines*, Presses universitaires de Rennes, 2019.

BRUCE, T., HILLMANN, D., *The Continuum of Metadata Quality: Defining, Expressing, Exploiting*, 2004. [En ligne],
https://www.researchgate.net/publication/247818823_The_Continuum_of_Metadata_Quality_Defining_Expressing_Exploiting

05.

GESTION DE L'ATELIER DE NUMÉRISATION

Si ce guide s'est jusqu'à présent concentré sur les documents numériques, il convient de s'intéresser également aux originaux numérisés. Ces objets précieux et souvent fragiles dans le cas de manuscrits anciens par exemple, demandent de prêter une attention particulière à leur manipulation, mais également de considérer l'environnement qui les entoure afin de limiter leur détérioration. C'est tout l'enjeu de la conservation préventive, définie par l'ICCROM¹ comme « l'ensemble des mesures et actions ayant pour objectif d'éviter et de minimiser les détériorations ou pertes à venir d'un objet d'art ou d'une collection patrimoniale²». Ces mesures et actions sont indirectes : elles n'interfèrent pas avec les matériaux et structures des biens et ne modifient pas leur apparence.

5.1 Préserver les collections patrimoniales et documentaires

Il existe dix facteurs principaux de détérioration du patrimoine³ :

- La température,
- L'humidité,
- La lumière,
- La pollution atmosphérique,
- La poussière, les moisissures,
- Les insectes et animaux,
- Le facteur humain, comprenant le vol et le vandalisme,
- Les sinistres, regroupant les incendies, les inondations et les conflits armés.

¹ Le Centre International d'Etudes pour la Conservation et la Restauration des biens Culturels (ICCROM) est une organisation intergouvernementale qui travaille au service de ses États membres pour promouvoir la conservation de toutes les formes de patrimoine culturel, dans toutes les régions du monde.

² ICCROM, *Preventive conservation*, n.d.
<https://www.iccrom.org/fr/section/conservation-preventive>

³ Canadian Conservation Institute, *Agents of deterioration*, 2017
<https://www.canada.ca/en/conservation-institute/services/agents-deterioration.html>

Les matériaux organiques et cellulotiques comme le papier, le parchemin ou le cuir sont parmi les plus sensibles à l'humidité relative⁴, car ils sont hygroscopiques, c'est-à-dire susceptibles d'absorber et de désorber l'humidité. Une variation trop importante de l'humidité relative peut notamment provoquer des variations dimensionnelles qui elles-mêmes entraînent des fendillements, des cassures, des soulèvements, des déformations⁵, etc.

Le tableau de la page suivante⁶ reprend les mesures environnementales et les variations autorisées pour la conservation optimale des matériaux les plus communément présents dans les collections des services d'archives et de bibliothèques.

La difficulté de la conservation préventive réside dans l'hétérogénéité des objets et documents rassemblés au sein d'un même espace : les conditions optimales pour la conservation du papier ne sont pas les mêmes que pour une photographie ou un cuir. **Pour autant, ces chiffres ne sont que des indicateurs, car plus que de s'attacher aux chiffres, il s'agit essentiellement de créer l'environnement le plus stable possible pour les collections patrimoniales tout en maintenant le confort des équipes.** L'excellent *Vade-mecum* de la conservation préventive du **Centre de Recherche et de Restauration des Musées de France (C2RMF)** regroupe les recommandations pour maîtriser aux mieux les conditions environnementales des collections et réagir en cas de constat d'anomalies⁷, dont voici les plus importantes :

- Mesurer l'humidité relative, la température et la luminosité de l'environnement par la mise en place de thermohygrographes et de luxmètres.
- Dresser une carte climatique du bâtiment en notant les mesures sur un plan.

- Repérer les zones à risques
- Identifier les sources des problèmes,
- Réduire au maximum les variations de température et d'humidité
- Si nécessaire, revoir la distribution des objets en fonction des zones climatiques
- Effectuer un contrôle régulier et entretenir les appareils de mesure et le matériel spécifique
- Connaître les zones les plus stables pour y mettre les collections les plus fragiles
- Modifier les conditions climatiques de certaines zones si nécessaire afin de savoir quand et comment déplacer les collections.
- Concevoir un plan d'urgence et un plan d'évacuation en cas de sinistre.

4 L'humidité est la teneur en vapeur d'eau de l'air, et sa teneur dans l'atmosphère ambiante est mesurée en humidité relative, qui exprime « la quantité de vapeur d'eau contenue dans un volume d'air donné par rapport au maximum qu'il pourrait contenir à une température et une pression données ».

Source : Etienne Fehau, Nathalie Le Dantec, *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, p.10.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

5 Jean-Paul Oddos (dir.), *La conservation : principes et réalités*, Paris : Éd. du Cercle de la Librairie, Coll. Bibliothèques, 1995, 405 p.

6 Ca tableau synthétise les informations trouvées dans les documents suivants :

Conseil Canadien des Archives, *Manuel de conservation des documents d'archives*, chapitre 3 - Environnement, Conseil canadien des archives, 2003, 18 p.
http://www.cdncouncilarchives.ca/RBch3_fr.pdf

Pierre Diaz Pedregal, *Climat des magasins d'archives : objectifs, moyens et méthodes - Petit manuel de climatologie appliquée à la conception des bâtiments d'archives*, Direction des Archives de France, 2009, 159 p.
<https://francearchives.fr/file/47cc986ffbb4b6b02960a1bcbd2b4a6a890cc098/Manuel%20de%20climatologie.pdf>

Etienne Fehau, Nathalie Le Dantec, *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, p.10.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

7 Etienne Fehau, Nathalie Le Dantec, *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, p.10.
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

TYPE	TEMPÉRATURE	HUMIDITÉ RELATIVE
Papier	<ul style="list-style-type: none"> • 17 °C à 18,3 °C Variation : 2°C sur 24h 	<ul style="list-style-type: none"> • 45-55% Variation : 3% sur 24h
Parchemin et cuir	<ul style="list-style-type: none"> • 17 °C à 18,3 °C Variation : 2°C sur 24h 	<ul style="list-style-type: none"> • 45-55% Variation : 3% sur 24h
Textile	<ul style="list-style-type: none"> • 17 °C à 18,3 °C Variation : 2°C sur 24h 	<ul style="list-style-type: none"> • 45-55% Variation : 3% sur 24h
Photographie		
Épreuves en noir et blanc	<ul style="list-style-type: none"> • Inférieure à 18°C 	<ul style="list-style-type: none"> • 30 - 50 %
Négatifs sur verre	<ul style="list-style-type: none"> • 18°C 	<ul style="list-style-type: none"> • 30 - 40 %
Négatifs noirs et blancs en gélatine argentique sur base de polyester	<ul style="list-style-type: none"> • 21°C +/- 2° sur 24h 	<ul style="list-style-type: none"> • 20 - 50 %
Négatifs en acétate de cellulose et en nitrate de cellulose	<ul style="list-style-type: none"> • 18°C 	<ul style="list-style-type: none"> • 35 - 60 %
Négatifs noirs et blancs en gélatine argentique sur base de triacétate	<ul style="list-style-type: none"> • 2°C 	<ul style="list-style-type: none"> • 20 - 50 %
Films couleur	<ul style="list-style-type: none"> • 2°C 	<ul style="list-style-type: none"> • 20 - 30 %
Épreuves couleur	<ul style="list-style-type: none"> • Inférieure à 2°C 	<ul style="list-style-type: none"> • 30 - 40 %
Microfilms		
Négatifs noirs et blancs en gélatine argentique sur base de polyester,	<ul style="list-style-type: none"> • Maximum 21°C 	<ul style="list-style-type: none"> • 20-50%
Films argentiques traités		
Films vésiculaires sur base de polyester		
Image argentique traitée par blanchiment sur base de polyester		

tab. 4 Température et humidité relative recommandées pour la conservation du papier, des textiles et du cuir

5.2 La gestion de la lumière

La lumière est la petite partie des ondes électromagnétiques qui est détectée par notre système visuel. On l'appelle rayonnement visible pour indiquer qu'il existe d'autres rayonnements qui sont invisibles à notre système visuel, mais qui accompagnent, dans des proportions diverses, le rayonnement visible : l'ultraviolet et l'infrarouge. En effet, toutes les sources de lumière, du soleil aux tubes fluorescents en passant par les lampes halogènes, émettent une très grande quantité de radiations, dans le domaine du visible comme de l'invisible. Jean-Jacques Ezrati a listé la composition spectrale des sources lumineuses les plus communes, repris dans le tableau ci-dessous.

Il s'agit essentiellement de créer l'environnement le plus stable possible pour les collections patrimoniales tout en préservant le confort des équipes.

SOURCE LUMINEUSE	RAYONNEMENT ULTRAVIOLET	RAYONNEMENT VISIBLE	RAYONNEMENT INFRAROUGE	RAYONNEMENT GLOBAL
Lumière d'un jour moyen	6%	44%	50%	100%
Lampe tungstène-halogène	1%	9.5%	90%	100%
Tube fluorescent 3000 K	1%	89%	10%	100%
Tube fluorescent 5000 K	2%	88%	10%	100%
Diode électroluminescente	0%	100%	0%	100%

tab. 5 Les différents types d'éclairants et leurs rayonnements
Source : Jean-Jacques Ezrati, *Éclairage d'exposition, musée et autres espaces*, Paris : Eyrolles, 2015, p.97

Ces rayonnements exercent des actions photochimiques et/ou thermiques sur les matériaux organiques, entraînant une dégradation irréversible de l'œuvre. Les ultraviolets ont une action photochimique et sont responsables de changements de couleur, du brunissement du papier et de l'affaiblissement mécanique général des œuvres. Les infrarouges produisant de la chaleur ont, quant à eux, un effet direct de ramollissement, d'assèchement ou d'inflammation sur les matériaux organiques.

Plus largement, ils accélèrent leur dégradation chimique. Les œuvres sensibles à la lumière se dégradent à chaque fois qu'elles y sont exposées. Cette dégradation est cumulative et irréversible. L'objectif est donc de supprimer le rayonnement ultraviolet, réduire le rayonnement infrarouge, et contrôler le rayonnement visible. Pour cela, la pose de stores et de filtres anti-UV sur les vitres, le contrôle par luxmètre et la pose de diodes électroluminescentes constituent des solutions intéressantes.

En somme, la protection des objets patrimoniaux face à la lumière **consiste à limiter la quantité de lumière et à en diminuer l'intensité ou la durée d'exposition.**

Parallèlement, la gestion de la lumière est un point particulièrement sensible pour les ateliers de numérisation. Pour rappel, l'enjeu de la numérisation est de restituer le plus fidèlement possible le document original. Or, une lumière trop intense viendrait altérer la perception des couleurs et donc la fiabilité de la reproduction numérique.

Maîtriser la lumière sous-entend **qu'aucune autre source ne parasite l'éclairage mis en place, aussi est-il indispensable de travailler dans un environnement imperméable à la lumière du jour.**

Dans le cas contraire, une machine qui aurait été calibrée durant l'hiver surexposerait toutes les images lors de périodes plus lumineuses, provoquant donc une perte irréversible d'information.

D'autres sources lumineuses – éclairage d'appoint pour la lecture de documents administratifs ou pour éviter des accidents, écrans d'ordinateurs – peuvent être ajoutées dans la pièce de numérisation à condition qu'elles n'aient pas d'influence sur la lumière réfléchie par l'objet.

Il est donc indispensable que ces sources ne soient pas dirigées vers l'objet et que leur puissance soit très faible – maximum 32 lux sur l'objet.

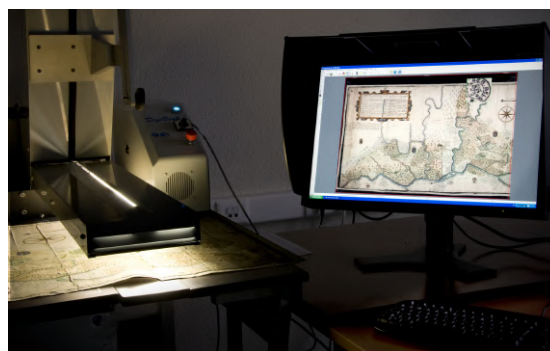


fig. 9 Atelier de numérisation

Archives de l'État belge

POUR ALLER PLUS LOIN

L'**ICCROM**, organisation intergouvernementale au service de ses États membres, promeut la conservation du patrimoine culturel sous toutes ses formes et dans le monde entier. À cet égard, plusieurs outils sont mis à dispositions des institutions culturelles pour les soutenir dans cette démarche :

- Un guide de gestion des risques pour la patrimoine culturel :
<https://www.iccrom.org/publication/guide-risk-management>
- La méthode ABC, qui consiste en une méthodologie pour identifier, évaluer et corriger les risques et facteurs participant à la dégradation des collections patrimoniales :
<https://www.iccrom.org/publication/abc-method-risk-management-approach-preservation-cultural-heritagf>

L'**Institut Canadien de Conservation (ICC)** propose un outil de référence rapide fournissant aux professionnels du patrimoine des conseils pratiques sur la façon de protéger les collections des agents de détérioration. Ces recommandations sont divisées en trois volets : le bâtiment et les installations, l'équipement et le matériel et enfin les procédures.

<https://www.canada.ca/fr/institut-conservation/services/conservation-preventive/plan-preservation-collections-patrimoniales.html>

Le **Centre de Recherche et de Restauration des Musées de France (C2RMF)** mène et publie depuis sa création des recherches sur la conservation et la restauration du patrimoine de même que sur la gestion des archives et l'usage de nouvelles technologies à cet effet. Sur son site, le centre propose une méthodologie pour accompagner les musées et institutions culturelles dans la rédaction de leur plan de conservation et sauvegarde.

<https://c2rmf.fr/conserver/fiches-techniques>

En Belgique, l'**Institut Royal du Patrimoine Artistique (IRPA)** compte en son sein une cellule dédiée à la recherche en conservation préventive. L'institut propose un service de conseils et d'analyse aux institutions culturelles et scientifiques qui souhaiteraient améliorer leur approche de la conservation préventive.

<https://www.kikirpa.be/fr/conservation-restauration/cellule-conservation-pr%C3%A9ventive>

EZRATI, J.-J. *Éclairage d'exposition, musée et autres espaces*, Paris : Eyrolles, 2015.

CONSEIL CANADIEN DES ARCHIVES. *Manuel de conservation des documents d'archives. Chapitre 3 – Environnement.* 2003.

http://www.cdncouncilarchives.ca/RBch3_fr.pdf

FEHAU E., LE DANTEC N., *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013.

https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf



06.

FICHES MÉTHODOLOGIQUES POUR ASSURER LA QUALITÉ D'UN PROJET DE NUMÉRISATION

*Les fiches présentées ci-dessous suivent l'ordre chronologique de la chaîne de la numérisation et constituent des **synthèses des éléments à prendre en considération lors des différentes étapes de la numérisation.***

*Elles sont chacune composées de trois catégories : les **concepts-clés**, qui reprennent brièvement les éléments théoriques à connaître sur l'étape développée, les **recommandations** sous forme de listes, et enfin les **outils** pour mettre en application ces recommandations.*

AVANT

Définir la stratégie de gestion de données

Définir la politique de numérisation

DÉFINITION INTELLECTUELLE ET STRATÉGIQUE DU PROJET DE NUMÉRISATION

Définir ses référentiels

Instaurer et appliquer une démarche qualité

Définir la qualité des livrables

Quand et comment faire intervenir le contrôle qualité

Contrôler la qualité en cas d'externalisation

DÉFINITION DU VOLET TECHNIQUE

Choisir le matériel de numérisation

Choisir les éclairants

Étalonner la chaîne de numérisation

PRODUCTION

Règles générales de prises de vue

Créer et remplir le fichier de récolement des métadonnées

Enregistrer et nommer les fichiers

POST-PRODUCTION

Retoucher les fichiers images

Océriser les documents textes

Encapsuler les métadonnées

Dernier contrôle qualité

DIFFUSION & CONSERVATION

Choisir son système de gestion de données numérique

01/ DÉFINIR SES RÉFÉRENTIELS

CONCEPT-CLÉ

Un référentiel est un **ensemble cohérent ou outillé de données ou d'informations partagées par une communauté d'acteurs.**

Il est caractérisé par sa stabilité, sa fiabilité, son unité de sens et son interopérabilité. Ce type de document s'inscrit dans le développement massif des échanges d'informations et de la communication et découle de la nécessité d'un langage commun pour ces échanges. L'usage de référentiels s'inscrit dans une perspective dynamique servant par exemple à identifier, analyser, classer, comparer ou mettre à jour des processus et des produits.

Dans le cadre d'un projet de numérisation, trois typologies de référentiels se démarquent :

- **Les normes.**
- **Les référentiels documentaires internes.**
- **Les référentiels méthodologiques.**

Les normes

Les normes sont des définitions standardisées et documentées qui centralisent les bonnes pratiques et procédés appliqués à des domaines variés¹. Au niveau international l'ISO (International Organization for Standardisation) fait référence en la matière. Il existe également des organismes de normalisation dédiés à des domaines spécifiques : c'est le cas par exemple du *World Wide Web Consortium (W3C)* ou de l'*Organization for the Advancement of Structured Information Standards (OASIS)* qui se concentrent sur les normes en matière de Technologies de l'Information.

¹ Si le cahier des charges d'une norme est entièrement adopté par un organisme, il est possible de demander une certification qui souligne le sérieux, les compétences techniques et la fiabilité de l'organisme. Si cette démarche est intéressante à bien des égards, elle demeure néanmoins optionnelle. En Belgique, BELAC est depuis 2006 l'unique organisme en charge de ces homologations.

Les référentiels documentaires

Les référentiels documentaires internes se manifestent le plus souvent sous la forme d'audit. Ce sont des documents plus ou moins élaborés propres à une organisation regroupant les processus, ressources et informations importantes à connaître sur un sujet donné.

Ce type de document tend à être de plus en plus standardisé, notamment grâce à la norme ISO 19001 qui offre un cadre rigoureux et complet pour l'élaboration de ces documents. Sur son site, la BnF met également à disposition un schéma méthodologique pour l'élaboration des audits.

Les référentiels méthodologiques

Enfin, les référentiels méthodologiques sont comme leur nom l'indique des corpus de méthodes et de bonne pratique sur un sujet donné.

Ces documents évoluent par version et peuvent être de toute nature : généralistes ou spécifique à un domaine ou processus, interne ou non. Le plus souvent, ce genre de document doit être adapté aux spécificités de l'organisation ou de l'institution qui l'utilise.

RECOMMANDATIONS

- ➔ **ANCTIL, M.-C, LEGENDRE, M., MÜLLER, T ., MAILLET, D., BROSSEAU, K., RENAUD, L.,** *Recueil de règles de numérisation*, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014, [En ligne], <https://numerique.banq.qc.ca/patrimoine/details/52327/2426216>
- ➔ **BIZINGRE, J., PAUMIER, J., RIVIÈRE, P.,** *Les référentiels du système d'information*, Paris : Dunod, 2013.
- ➔ **BURESI, C., et CÉDELLE-JOUBERT, L.,** *Conduire un projet de numérisation*, Paris : Tec et Doc, 2002, 326 p.
- ➔ **CLAERR, T., WESTEEL, I., et MELOT, M.,** *Manuel de la numérisation*, Paris : Éd. du Cercle de la Librairie, 2011. 317 p.
- ➔ **CLAERR, T., et WESTEEL, I.,** *Numériser et mettre en ligne*, Villeurbanne : Presses de l'enssib, 2010, [En ligne], <http://books.openedition.org/pressesenssib/414>

- ➔ **CLAERR, T., WESTEEL, I., ORY-LAVOLÉE, B., GEORGES, N.,** *Manuel de constitution de bibliothèques numériques*, Paris : Éd. du Cercle de la Librairie, 2013. 407 p.
- ➔ **DIRECTION DES ARCHIVES DE FRANCE. MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION - COMITÉ DE PILOTAGE NUMÉRISATION - MISSION DE LA RECHERCHE ET DE LA TECHNOLOGIE (DDAI),** *Écrire un cahier des charges de numérisation du patrimoine. Guide technique. Documents reliés- manuscrits- plans- dessins- photographies- microformes*, Paris : Archives de France, 2008, 62p., [En ligne]
https://francearchives.fr/file/bf50d8fa5f554586dbf18fdc862d25970a1da0a7/static_4132.pdf
- ➔ **DIRECTION INTERMINISTÉRIELLE DU NUMÉRIQUE ET DU SYSTÈME D'INFORMATION ET DE COMMUNICATION DE L'ETAT,** *Référentiel général d'interopérabilité*, 2015, version 2, 84 p., [En ligne]
https://www.numerique.gouv.fr/uploads/Referentiel_General_Interoperabilite_V2.pdf
- ➔ **FEHAU E., LE DANTEC N.,** *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, 46 p., [En ligne],
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf
- ➔ **FORCE 11,** *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0*, n.d., [En ligne],
<https://www.force11.org/group/fairgroup/fairprinciples>
- ➔ **PLAN D'EXPLOITATION ET DE PRÉSERVATION DES PATRIMOINES (PEP'S).** *Normes et lignes directrices techniques et organisationnelles pour les initiatives de numérisation des patrimoines culturels soutenues par la Communauté française, ministère de la communauté française*, Bruxelles : Secrétariat général – Préservation et exploitation des patrimoines, 2009, 61 p., [En ligne],
http://www.peps.cfwb.be/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/sites/numpat/upload/numpat_super_editor/numpat_editor/documents/CFWB/Normes.pdf&hash=a2d0e740e46e34b8c3475c604c3e96edbb8a8712
- ➔ **RIEGER, T.,** *Federal Agencies Digital Guidelines Initiative. Technical Guidelines for Digitizing Cultural Heritage Material. Creation of Raster Image File*, 2016, 99 p., [En ligne],
http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf
- ➔ **VAN DORMOLEN, H.,** *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012, [En ligne],
http://www.imagingetc.com/images/Resources/Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

ISO 9000 :2015	Systèmes de management de la qualité <i>Principes essentiels et vocabulaire</i>
ISO 9001 :2015	Systèmes de management de la qualité <i>Exigences</i>
ISO 9004 :2018	Qualité d'un organisme <i>Lignes directrices pour obtenir des performances</i>
ISO 19011 :2018	Lignes directrices pour l'audit des systèmes de management
ISO 31000	<i>Risk Management</i>
ISO 15489-1 :2016	Information et documentation – Gestion des documents d'activité <i>Partie 1 : Concepts et principes</i>
ISO/IEC 27001:2013	Technologies de l'information – Techniques de sécurité – Systèmes de management de la sécurité de l'information <i>Exigences</i>
NF Z42-026	Définition et spécifications des prestations de numérisation fidèle de documents sur support papier et contrôle de ces prestations
ISO 2859-1	Règles d'échantillonnage pour les contrôles par attributs <i>Partie 1 : procédures d'échantillonnage pour les contrôles lot par lot, indexés d'après le niveau de qualité acceptable</i>
FD Z42-017: 2009	Imagerie électronique – Numérisation de documents – Sous-traitance d'opération de numérisation <i>Guide pour l'élaboration d'un cahier des charges techniques</i>
ISO 12641-1 :2016	Technologie graphique – Échange de données numériques de préimpression <i>Cibles de couleur pour étalonnage à l'entrée du scanner</i>
ISO 19264-1 :2021	Photography – Archiving systems – Imaging systems quality analysis <i>Part 1: Reflective originals</i>
ISO 7589 :2002	Photography – Illuminants for sensitometry <i>Specifications for daylight, incandescent tungsten and printer</i>
ISO 14524 :2009	Photography – Electronic still-picture cameras <i>Methods for measuring opto-electronic conversion functions (OECFs)</i>
ISO 21550 : 2004	Photography – Electronic scanners for photographic images <i>Dynamic range measurements</i>
ISO 25964-2:2013	Information et documentation <i>Thésaurus et interopérabilité avec d'autres vocabulaires</i>

02 / INSTAURER ET APPLIQUER UNE DÉMARCHE QUALITÉ

CONCEPT-CLÉ

La démarche qualité ou le management de la qualité est un concept managérial pour mieux appréhender et analyser les processus de travail et leurs résultats.

Ce concept est basé sur la méthode de gestion de qualité PCDA – Plan-Do-Change-Act – visant à créer un cercle vertueux œuvrant constamment à l'amélioration de la qualité des projets et des institutions.

Héritière de cette méthode, la famille de norme ISO 9000 propose un cadre méthodologique généraliste pour améliorer les processus de travail et leurs résultats :

Objectifs et avantages

- ⇒ Augmenter la performance individuelle de l'atelier de numérisation par l'amélioration de son organisation.
- ⇒ Mieux prendre en compte la demande des **utilisateurs** et leur proposer un contenu en adéquation avec leurs attentes.
- ⇒ Inscrire l'institution dans la **transition numérique européenne** et partager une culture organisationnelle globale.
- ⇒ Améliorer la **qualité des livrables**.

RECOMMANDATIONS

- ➔ Élaborer une stratégie de gestion de données.
- ➔ Établir une politique de numérisation.
- ➔ Pour chaque projet de numérisation, rédiger un cahier des charges.
- ➔ Créer un **workflow** du processus de numérisation.
- ➔ Définir précisément le rôle et les tâches de chaque membre de l'équipe.
- ➔ Utiliser un outil de suivi de projet commun à toute l'équipe pour faciliter la communication.
- ➔ En cours de projet, **consigner les erreurs, anomalies et dysfonctionnements et la réponse apportés à ces difficultés**.
- ➔ À la fin du chantier de numérisation, **évaluer avec l'ensemble de l'équipe le projet** en lui-même et en comparaison aux projets de numérisation passés : certaines mêmes erreurs ont-elles été rencontrées ? La réponse à ces dysfonctionnements a-t-elle été adaptée ? Les résultats escomptés ont-ils été atteints ?
- ➔ **Adapter les documents stratégiques et le workflow si nécessaire en vue d'améliorer les projets suivants**.

OUTILS

- **Norme ISO 9000 :2015 - Système de management de la qualité**

Principes essentiels et vocabulaire

Exigences

- **Norme ISO 9004 :2018 - Management de la qualité**

Qualité d'un organisme – Lignes directrices pour obtenir des performances durables

- **Outils de suivi de projet : Projector et Redmine**

Les outils open-source **Projector** et **Redmine** proposent de nombreuses fonctionnalités pour le suivi des tâches, la visualisation des ressources disponibles pour un projet ou pour la gestion du planning au sein d'une équipe.

POUR ALLER PLUS LOIN

BAROUCH, G., *Révolution des pratiques managériales : le changement par la qualité*, La Plaine Saint-Denis, Afnor Éditions, 2013.

GIESEN, E., *Démarche qualité et norme ISO 9001 : Une culture managériale appliquée à la recherche*, Marseille : IRD Éditions, 2018, 180 p.

CARBONE F., OOSTERBEEK L., COSTA C., FERREIRA AM., « Extending and adapting the concept of quality management for museums and cultural heritage attractions: A comparative study of southern European cultural heritage managers' perceptions », *In: Tourism Management Perspective*, vol.35, juillet 2020.

03 / DÉFINIR LA QUALITÉ DES LIVRABLES

CONCEPT-CLÉ

La **qualité** peut être définie comme l'ensemble des propriétés et caractéristiques d'un produit, processus ou service qui lui confèrent son aptitude à satisfaire les besoins exprimés ou implicites.

La **sous qualité ou non-qualité** correspond à un document numérisé ne répondant pas aux exigences minimales attendues, c'est à dire qu'il ne restitue pas la lisibilité ou l'intégrité esthétique du document original. À l'inverse, la **sur qualité** correspond à l'excès d'optimisation de la qualité d'un produit qui engendre l'utilisation de ressources et de frais trop importants pour une amélioration du produit sans réelle plus-value.

Enfin, le **seuil de qualité minimale ou acceptable** est la marge d'erreur autorisée lors de la production d'un projet de numérisation. Ces erreurs doivent être mineures et ne pas entraver la lisibilité, l'intégrité esthétique et l'exploitation du document.

Objectifs

- ⇒ La qualité se définit relativement à l'exploitation prévue des documents numérisés (conservation / diffusion / valorisation).
 - ⇒ Trois jalons sont à identifier lors de l'élaboration du projet : la non-qualité, la sur-qualité et le seuil de qualité acceptable. Pour chaque critère défini comme un
- élément clé de l'analyse de la qualité d'une production, il est nécessaire de définir des seuils au-dessus ou au-dessous desquels les images seront déclarées conformes ou non-conformes.
- ⇒ Ces seuils peuvent être basés sur une ou plusieurs valeurs pour chaque critère.

RECOMMANDATIONS

EXEMPLES D'ERREURES MINEURES

- ➔ Absence du fichier de récolement finalisé par le prestataire.
- ➔ Non-conformité des supports et des formats de fichiers numériques.
- ➔ Non-conformité du nommage appliqué aux supports, dossiers, fichiers et images.
- ➔ Non-conformité des répertoires et des structures de fichiers.
- ➔ Incohérence dans le contenu des champs principaux du fichier de récolement.
- ➔ Absence d'indexation demandée.
- ➔ Non-conformité à l'indexation principale aux règles fournies.
- ➔ Numérisation incomplète (documents non reproduits).

EXEMPLES D'ERREURES MAJEURES

- ➔ Incohérence dans le contenu des champs secondaires du fichier de récolement.
- ➔ Non-conformité de la saisie de l'indexation.
- ➔ Non-respect de l'ordre des vues.
- ➔ Mauvais cadrage sans incidence sur la lecture, et n'entraînant pas de perte d'information.
- ➔ Dérive mineure, en chromie ou/et en densité par rapport aux fichiers de référence.
- ➔ Mauvaise orientation des vues.
- ➔ Fidélité au document original altérée (problème de répartition lumineuse, de planéité, de perspective, de netteté, de rayures).
- ➔ Utilisation de vitres sales pour aplatir un document par exemple.

04 / QUAND ET COMMENT FAIRE INTERVENIR LE CONTRÔLE QUALITÉ

CONCEPT-CLÉ

Il existe **trois types de contrôle qualité** à faire intervenir durant la phase de numérisation :

- Le **contrôle visuel**, qui permet de vérifier la lisibilité et l'esthétisme de l'image.
Exemples : marges, couleur de fond, flou, etc.
- Le **contrôle technique**, qui consiste en la vérification des éléments techniques des contenus numérisés.
Exemples : résolution/définition, formats de fichiers, métadonnées (EXIF, modèle IPTC XMP ou autres), présence de profil colorimétrique, analyse des mires, etc.
- Le **contrôle d'intégrité ou de cohérence**, qui porte sur la vérification du respect des modèles qui ont un impact sur l'intégration des images dans l'outil de gestion des documents numériques.
Exemples : arborescence, nom de fichier, nom de répertoire, complétude, etc.).

Théoriquement, chaque étape du processus de numérisation peut faire l'objet d'un contrôle qualité : selon le résultat, le processus peut être poursuivi ou repris partiellement, voir totalement.

De façon pragmatique, **plus la détection de la non-conformité est précoce, plus le coût de la reprise est faible**. Un produit non conforme décelé par l'utilisateur final coûte au minimum le double d'un produit fabriqué convenablement en première instance.

Le coût de la reprise s'avère parfois bien plus élevé en réalité. Le produit non conforme découvert pendant la production représente un surcoût non négligeable, mais moindre par rapport à une reprise globale. La qualité trouve donc ici une justification financière tout à fait logique.

RECOMMANDATIONS

ÉTAPE	PARAMÈTRES À CONTRÔLER	OUTILS CONSEILLÉS
Préparation de l'environnement	<ul style="list-style-type: none"> • Température • Taux d'humidité relative • Luminosité 	<ul style="list-style-type: none"> • Thermo-hygromètres • Luxmètres
Choix des éclairants	<ul style="list-style-type: none"> • Disposition et inclinaison des éclairants autour du matériel de prise de vue 	
Sélection et préparation des documents	<ul style="list-style-type: none"> • État des document originaux • Éléments externes à enlever pour la numérisation (trombones, papiers volants, marque-pages, etc.) 	
Création et remplissage du fichier de récolement des métadonnées descriptives	<ul style="list-style-type: none"> • Cohérence entre les données entrées et le standard utilisé • Complétude et exactitude des métadonnées entrées • Cohérence du vocabulaire utilisé 	<ul style="list-style-type: none"> • OpenRefine
Étalonnage de la chaîne	<ul style="list-style-type: none"> • État du matériel : stabilité, propreté de l'objectif • Impact de la luminosité sur le matériel de prise de vue • Espace colorimétrique utilisé • Cohérence des profils ICC appliqués aux périphériques 	<ul style="list-style-type: none"> • AutoSFR • OpenDice • Universal Test Target (UTT)
Pré-production & Production	<ul style="list-style-type: none"> • Complétude • Exposition • Contraste • Couleur • Mise au point • Résolution • Parallélisme • Flare • Vignettage • Bruit • Artéfacts 	
Enregistrement et nommage des fichiers	<ul style="list-style-type: none"> • Cohérence entre les noms des fichiers et le plan de nommage de l'institution 	<ul style="list-style-type: none"> • File Renamer Basic • BulkRename

ÉTAPE	PARAMÈTRES À CONTRÔLER	OUTILS CONSEILLÉS
Retouche des documents image	<ul style="list-style-type: none"> ● Cohérence entre les profils ICC du matériel de prise de vue et le programme ● Présence et exactitude des métadonnées techniques ● Caractéristiques techniques des images 	<ul style="list-style-type: none"> ● Photoshop ● Lightroom ● Gimp
Océrisation des documents texte	<ul style="list-style-type: none"> ● Omissions ● Problèmes de classification du logiciel d'océrisation ● Problèmes de segmentation 	<ul style="list-style-type: none"> ● ABBY Fine Reader ● Tesseract ● Layout Evaluation
Encapsulation des métadonnées techniques et administratives	<ul style="list-style-type: none"> ● Complétude des métadonnées présentes : descriptives / administratives / techniques ● Exactitude des chemins entre les différents fichiers de métadonnées 	<ul style="list-style-type: none"> ● Standard METS
Dernier contrôle qualité	<ul style="list-style-type: none"> ● Cohérence entre les critères décrits dans le cahier des charge et les résultats obtenus : ● Conformité des formats ● Conformité du nommage des fichiers ● Complétude des images ● Caractéristiques techniques des images ● Présence et exactitude des métadonnées descriptives, techniques et administratives ● Complétude et exactitude du fichier METS 	

05 / CONTRÔLER LA QUALITÉ EN CAS D'EXTERNALISATION

CONCEPT-CLÉ

L'**externalisation** de la production désigne **la possibilité pour un organisme de confier l'ensemble ou une partie du processus de numérisation à un prestataire extérieur**. L'externalisation permet de répondre à un manque de moyens techniques ou de gérer un flux important de documents à numériser. Elle peut être également utilisée comme stratégie pour maîtriser les coûts liés à une opération ou comme option pour jouir d'une plus grande expertise. Pour autant, l'externalisation demande des démarches supplémentaires afin de ne pas perdre le contrôle de la chaîne de numérisation et préserver la qualité des livrables.

L'ampleur et la nature des tâches à confier au prestataire s'élaborent dès la rédaction du cahier des charges. Doivent également être précisément définies les modalités de contrôle.

RECOMMANDATIONS

- ➔ **Déterminer le périmètre d'action de l'ensemble des collaborateurs** : type de tâches, déroulé de la prestation, délais.
- ➔ **Définir les conditions générales d'exécution** (normes à suivre, référentiels méthodologiques, etc).
- ➔ **Définir les modalités de transport et de manipulation des documents originaux.**

- ➔ **Définir le mode de suivi** en cours de production et notamment le type de rapports et leur fréquence.
- ➔ **Définir les démarches à suivre en cas d'anomalies** en cours de production.
- ➔ **Fournir au prestataire le plan de nommage à suivre**, le standard de métadonnées à appliquer et les indexations à effectuer.
- ➔ **Partager la responsabilité du contrôle qualité des livrables.**
- ➔ **Définir des indicateurs de suivi de contrôle**, simples, fiables et pertinents.
- ➔ **Définir un pourcentage autorisé d'erreurs.**

Exemple : la Bibliothèque nationale de France

Dès ses premiers chantiers de numérisation à grande échelle, le **BnF** a choisi d'externaliser une partie de ses chantiers de numérisation, et donc de réviser sa politique de contrôle qualité pour garantir un dialogue aisé et transparent avec les prestataires externes.

La solution adoptée alors fut **l'élaboration d'une démarche qualité orientée autour du principe de fluidité** avec comme principes fondamentaux :

L'implication du prestataire dès l'élaboration du cahier des charges, et notamment de s'accorder sur les modalités de travail, la qualité du document numérique, les processus de la chaîne de numérisation, les indicateurs à surveiller, etc.

La mise en place d'un dialogue constant, notamment sous la forme de rapports hebdomadaires établis par le prestataire et indiquant le nombre de documents en entrée et en sortie de chaque atelier, ainsi que le

nombre de pages entrées et sorties de chaque atelier. Ces rapports permettent ainsi d'identifier les étapes problématiques, d'analyser les causes de ces engorgements et de corriger précisément la chaîne de numérisation en cours de production.

Le partage de la responsabilité du contrôle qualité.

Pour plus de détails consulter :

BELLIER, L., « Numérisation, pour une nouvelle approche de la qualité », *In: Bulletin des bibliothèques de France*, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB), 2014.

<https://bbf.enssib.fr/contributions/numerisation-pour-une-nouvelle-approche-de-la-qualite>

06 / CHOISIR LE MATÉRIEL DE NUMÉRISATION

RECOMMANDATIONS

L'offre des scanners pour la numérisation patrimoniale et documentaire n'a jamais été aussi importante. Devant la surenchère des constructeurs en matière d'innovations techniques, il est parfois difficile de dégager les critères de sélection fondamentaux du matériel de numérisation.

Choisir le matériel de numérisation demande avant tout de se pencher sur :

- ➔ Le **type** de documents à numériser.
- ➔ Leurs **formats**.
- ➔ La **quantité** de documents et leurs **préciosité**.
- ➔ Leur **état**, et la manière dont ils peuvent être manipulés en cours de processus.
- ➔ Le **budget** disponible.
- ➔ Le **degré de qualification** des équipes.
- ➔ Le **degré de complexité d'utilisation du matériel** et la formation éventuelle des équipes pour maîtriser l'équipement.
- ➔ L'**espace** pour entreposer les machines.
- ➔ Les **conditions environnementales** de l'atelier de numérisation.

Une fois ces éléments déterminés, le choix du matériel de prise de vue se base sur des **caractéristiques purement techniques** :

- ➔ La résolution, ou la précision de la capture de l'image.
- ➔ La profondeur d'analyse, ou le nombre de couleurs et de niveaux de gris que l'appareil est capable de produire.
- ➔ L'**interface**, la présence de ports USB.
- ➔ La **vitesse de numérisation**.

Les appareils de prises de vue adaptés et recommandés pour la numérisation patrimoniale et documentaire sont :

- ➔ Les **appareils photos**.
- ➔ Les **scanners à tambour**.
- ➔ Les **bancs de reproduction**.
- ➔ Les **scanners spécialisés dans la numérisation de reliure**.

Attention

- ➔ À force d'utilisation, ou suite à une erreur de manipulation, **les éléments d'un système peuvent se dégrader ou se dérégler** sans que ce ne soit forcément repérable à l'œil nu.
- ➔ C'est pourquoi il est important de **faire régulièrement un contrôle qualité des images produites**. La fréquence de ce type de contrôle dépend de l'intensité de l'usage des machines et des recommandations du constructeur. Ce type de contrôle permet en outre de valider les images produites entre deux tests et de participer à l'optimisation d'un éventuel travail d'impression.

07 / CHOISIR LES ÉCLAIRANTS

CONCEPT-CLÉ

Comme souligné, la lumière naturelle détériore les objets patrimoniaux. Le choix des éclairants, leur puissance et leur positionnement est donc essentiel pour optimiser la qualité du processus de numérisation.

La qualité de l'éclairage s'évalue par sa capacité à rendre perceptible l'ensemble du spectre des couleurs, qui est définie comme la capacité discriminatoire. Cette capacité se mesure en Indice de Rendu des Couleurs (IRC) grâce à un thermocolorimètre dédié à l'image numérique. La valeur maximale de l'IRC est de 100 et correspond à une lumière du jour dite idéale. Sa valeur minimale est de 0 et correspond à une lumière qui rendrait impossible toute distinction de couleur. **La valeur IRC minimale pour assurer une bonne reproduction des couleurs est de 95.**

Pour atteindre ce seuil et assurer l'uniformité de l'éclairage, **les éclairants doivent être de même qualité et de même puissance, être positionnés de part et d'autre du plan de capture, de manière symétrique, et être dirigés vers le centre de ce dernier selon un angle de 30 à 45 degrés**. Ce faisant, leur hauteur doit être au minimum égale à celle de l'objectif, ce dernier étant équipé d'un dispositif pare-soleil – pour éviter notamment l'effet de flare.

RECOMMANDATIONS

➔ **Le flash électronique** est la solution la plus à même de restituer correctement l'apparence d'un objet. Avec un IRC avoisinant les 98%, la lumière au flash dispose d'une puissance lumineuse améliorant les valeurs de contraste.

- ➔ Les **lampes aux halogénures métalliques** – HMI, MSR, GEMI, etc. – sont une alternative intéressante. Initialement dédiées au cinéma et à la télévision, elles produisent une lumière stable dont le spectre est large et complet. Calibrée entre 5 000 et 5 500 Kelvins, leur IRC avoisine les 95%.
- ➔ Les **types d'éclairage déconseillés pour la numérisation** sont les lampes à fluorescence dont l'IRC est trop faible et les LED qui ne permettent pas un rendu optimal des couleurs.
- ➔ Les **deux éclairants utilisés doivent être positionnés de part et d'autre du matériel de numérisation**. Dans le cas d'un document de grand format, il est préférable d'opter pour quatre sources lumineuses.
- ➔ La lumière doit être **tamisée et constante**. La luminosité maximale ne doit pas excéder les **32 lux**.
- ➔ **L'écran doit être protégé de la lumière du jour** ou des reflets directs des éclairants pour éviter la réverbération de la lumière sur l'équipement de numérisation.

08 / ÉTALONNER LA CHAÎNE DE NUMÉRISATION

CONCEPT-CLÉ

L'étalonnage consiste à paramétrer l'ensemble du matériel de prise de vue et ses périphériques selon les spécifications techniques exprimées dans le cahier des charges. Il permet de régler les paramètres suivants :

- La résolution
- La balance des blancs
- L'exposition
- La mise au point
- Le contraste
- Les couleurs et leur rendu

Une gestion des couleurs homogène et précise nécessite **l'utilisation de profils ICC fiables pour tous les périphériques couleur**. Sans le contrôle de ces profils, une image numérisée peut présenter des différences chromatiques d'un périphérique à l'autre, en raison d'une simple différence entre le scanner et le logiciel d'affichage graphique. Avec un profil fiable, le programme qui importe l'image peut rectifier les différences entre les périphériques et ainsi afficher les couleurs réelles d'une image numérisée. Il est donc essentiel – quel que soit le profil colorimétrique choisi – **d'aligner ces profils entre les différents périphériques et de les calibrer de la même manière**.

Attention



Chaque scanner a ses faiblesses dans le rendu des couleurs. Il est donc indispensable de **calibrer régulièrement le matériel de prise de vue en numérisant un document et en comparant les couleurs du fichier numérique et de l'original pour s'assurer de la fiabilité du rendu**. Une charte couleurs – ou mire couleurs – déposée à côté de l'objet scanné permet de mieux appréhender le rendu des couleurs.

OUTILS

Les normes et référentiels

- **Norme ISO 12641-1 :2016 - Technologie graphique**

Échange de données numériques de préimpression - Cibles de couleur pour étalonnage à l'entrée du scanners

- **Norme ISO 7589 :2002 - Photography**

Illuminants for sensitometry – Specifications for daylight, incandescent tungsten and printers

- **Norme ISO 14524 :2009 - Photography**

Electronic still-picture cameras – Methods for measuring opto-electronic conversion functions (OECFs)

- **Norme ISO 21550 : 2004 - Photography**

Electronic scanners for photographic images – Dynamic range measurements

- **Guide Metamorfoze**

http://www.imagingetc.com/images/Resources_Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

- **Guide FADGI**

http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

Les espaces colorimétriques et profils ICC

- **Adobe RVB 1998**

Espace colorimétrique privilégié pour la numérisation couleur composé de trois niveaux (Rouge-Vert-Bleu)

- **Adobe sRVB**

Espace colorimétrique alternatif pour la numérisation couleur composé de trois niveaux (Rouge-Vert-Bleu)

- **Gray Gamma 2.2**

Profil ICC privilégié pour les niveaux de gris. Permet de régler la balance des blancs et des gris.)

- **Kodac Gray Scale**

Profil ICC alternatif pour les niveaux de gris. Permet de régler la balance des blancs et des gris.

- **Digital Color Checker**

Profil ICC privilégié pour les numérisation couleur.

Les outils d'évaluation

- **Open DICE**

Logiciel libre de mesure et d'analyse des critères techniques des scanners. Ce logiciel de contrôle automatique s'appuie notamment sur plusieurs normes ISO pour analyser la qualité des images produites par les scanners et les composants techniques du matériel de prise d'image.

<http://www.digitizationguidelines.gov/guidelines/OpenDICE/autoSFR.pdf>

- **Auto SFR**

Programme libre développé pour aider les professionnels de l'imagerie à déterminer la résolution réelle des images et à définir la résolution adaptée aux documents à numériser selon leur type.

http://www.digitizationguidelines.gov/guidelines/OpenDICE/AutoSFR_manual.pdf

- **UTT**

Method for checking and controlling all the parameters of an image capture system, developed by the National Library of the Netherlands.

This standardised test target – available in formats ranging from DIN A4 to DIN A0 – makes it possible to validate parameters such as: resolution, contrast, white balance, gain modulation, uniformity of light on the object, noise, colour rendition, geometric distortion or the parallelism of a capture solution.

<http://universaltesttarget.com/about.php>

RECOMMANDATIONS

- ➔ Le **réglage des éclairants doit être réalisé en amont de l'étalonnage** afin de ne pas fausser ce dernier. La luminosité maximale autorisée est de 32 lux.
- ➔ Résolution préconisée :
 - **300 DPI** pour les documents compris entre le DIN5 et le DIN A2.
 - **400 DPI** pour les autres formats.
- ➔ Privilégier l'utilisation des **espaces colorimétrique Adobe RVB 1998** pour la numérisation couleur et **Gray Gamma 2.2** pour les images en niveaux de gris.
- ➔ Pour s'assurer de la qualité de reproduction des tonalités et teintes du document original, il est nécessaire de **numériser la charte de couleur référentielle dans les mêmes conditions que les documents reproduits.**
- ➔ Cette **mire doit être numérisée tous les jours** pour s'assurer que le paramétrage des machines reste conforme.

09 / RÈGLES GÉNÉRALES DE PRISE DE VUE

RECOMMANDATIONS

- ➔ Les documents doivent être complètement numérisés : **l'image doit refléter la totalité du document** et **comprendre toutes les pages** dans le cas de la numérisation d'un ouvrage.
- ➔ Pour certifier que le document est complet, **laisser une marge sur tout le pourtour du document de 0,25 à 1 cm**. Plus le document est grand, plus cette marge sera importante.
- ➔ Si les contours du document à reproduire sont irréguliers, laisser une bordure suffisamment large pour cadrer l'intégralité du document sur l'ensemble de ses côtés.
- ➔ Il est préconisé de **numériser les documents en couleurs**.
- ➔ Les documents doivent être **mis à plat autant que possible**, dans la limite de la contrainte du document, de son état et de sa préciosité.
- ➔ Cette mise à plat doit être **uniforme et homogène** pour éviter toute déformation du document.
- ➔ L'orientation du document doit **respecter le sens de lecture**. Si le format du document ne le permet pas, il devra être orienté dans ce sens lors de la retouche en post-production.

- ➔ Pour les documents contenant du texte imprimé, **la ligne de ligne de texte doit être horizontale avec une variation tolérée de 1,30 degré.**
- ➔ Privilégier **la rectitude du texte** plutôt que celle de la page pour éviter de compromettre l'étape de reconnaissance de caractères des documents (OCR).
- ➔ Placer **un fond de numérisation neutre - gris, noir ou blanc** – afin d'accroître le contraste et donc la lisibilité du document.
- ➔ Pour les documents reliés présentant des lacunes, **placer un cache en carton mat de couleur neutre** pour accroître sa lisibilité et ne pas compromettre l'OCR.
- ➔ Pour pouvoir rectifier les erreurs constatées en cours de procédure, **conserver le fichier RAW** afin de pouvoir le dématricer à nouveau et corriger l'anomalie.

10 / CRÉER ET REMPLIR LE FICHIER DE RÉCOLEMENT DES MÉTADONNÉES

CONCEPT-CLÉ

Le **fichier de récolement** est un fichier regroupant les **métadonnées descriptives du document** numérisé afin d'en connaître l'auteur, le volume, les dimensions, les caractéristiques de conservation, et plus largement, tous les éléments essentiels à connaître pour les opérateurs de numérisation en vue de sa diffusion et de son catalogage.

La définition de ces catégories dépend une nouvelle fois de **l'exploitation prévue des documents** numérisés, mais également de l'usage ou non d'un modèle de métadonnées standardisé – dont les catégories sont déjà définies.

Dans le cas d'une externalisation, le fichier de récolement doit être créé en amont et transmis au prestataire. Une fois rempli et remis, ce fichier devra faire l'objet d'un contrôle d'intégrité pour s'assurer de l'exactitude des informations créées.

RECOMMANDATIONS

Le fichier de récolement doit au minimum comprendre les informations suivantes :

- ➔ Les **catégories descriptives** des objets numérisés (identifiant unique, auteur, titre, date de création, sujet, dimensions, description, etc.).
- ➔ La **définition** de ces catégories.

- ➔ Le **degré d'obligation des catégories** de métadonnées (obligatoire / optionnel / recommandé).
- ➔ Le **type de données** –quantité, date, vocabulaire contrôlé, texte libre, etc.
- ➔ Les **normes de saisies** pour les textes libres et le document référentiel du standard de métadonnées suivi par l'institution.
- ➔ Le **nom de l'encodeur et les modifications apportées** afin de suivre les modifications et d'identifier les éventuels problèmes de fiabilité.

OUTIL

- **Open Refine**

Open Refine est un programme open-source permettant de nettoyer et de corriger un ensemble de données, leur apportant ainsi plus de cohérence. Il permet ainsi :

- *L'importation de données dans différents formats*
- *L'application de transformation de cellules*
- *Le traitement des cellules qui comprennent plusieurs valeurs*
- *La création de liens entre des ensembles de données*

<https://openrefine.org/documentation.html>

11 / ENREGISTRER ET NOMMER LES FICHIERS

CONCEPT-CLÉ

Tout projet de numérisation pose la question du nommage des fichiers qui seront produits à l'issue des opérations. Cette problématique entre en résonance avec le contexte du Web sémantique, dont l'un des principes fondamentaux est l'utilisation d'identifiants pérennes.

Un identifiant est dit pérenne lorsqu'il désigne de façon unique une ressource ainsi que l'ensemble des métadonnées qui lui sont associées, permettant de fait de la découvrir plus facilement sur le Web. Cet identifiant – comme par exemple l'URI, l'ARK ou le DOI – garantit également la réutilisation des ressources numériques et leur visibilité en offrant la possibilité de les lier à d'autres ressources et ainsi démultiplier les parcours de lecture ou chemins d'accès.

Face au nombre important de documents à gérer dans un projet de numérisation, il n'est pas envisageable de nommer les fichiers individuellement. **L'utilisation de programmes semi-automatisés permet la gestion de masse des identifiants de fichiers, et ce peu importe leur type, leur poids ou leur localisation.**

RECOMMANDATIONS

- ➔ Quel que soit le programme de gestion choisi, il est essentiel de **définir une convention de désignation des fichiers** afin de faciliter le travail de numérisation et diminuer le risque de perte d'informations.

➔ Il n'existe pas de façon parfaite de structurer et d'ordonner les éléments présents dans un nom de fichier. Il importe avant tout d'être **uniforme**. Voici quelques-unes des catégories les plus utilisées pour identifier des documents au sein des institutions culturelles et scientifiques :

- Identifiant alphanumérique
- Type de document
- Date de création du document (format de date AAAAMMJJ)
- Objet du document
- Destinataire dans le cas d'une demande de reproductions
- Version

➔ **Règle de forme à respecter :**

- Séparation des mots dans un champ majuscule ou underscore

➔ **Caractères proscrits :**

- Caractères accentués
- Signes de ponctuation
- Espace

➔ Il est préférable de **se conformer au plan de nommage en vigueur au sein de son institution** ou, s'il n'en existe pas encore, d'en définir un qui sera propre aux opérations de numérisation.

➔ **Garantir l'unicité des identifiants.**

➔ En cas de suppression ou de dépublication, en renseigner les circonstances et le cas échéant, **pointer vers une ressource de substitution.**

➔ Publier sa **politique de gestion des identifiants.**

OUTILS

● **File Renamer Basic**

*File Renamer Basic est un programme de renommage de fichiers pour **Windows**. Il peut appliquer un format de nom à tous les fichiers d'un même dossier et est capable d'éditer par lots les métadonnées techniques des fichiers images (EXIF).*

https://download.cnet.com/File-Renamer-Basic/3000-2248_4-10306538.html

- **Bulk Rename Utility**

*Bulk Rename Utility est un programme de renommage de fichiers pour **Windows**. Il propose les mêmes fonctionnalités que File Renamer Basic, mais a été créé spécialement pour supporter les dossiers contenant plusieurs milliers de fichiers.*

<https://www.bulkrenameutility.co.uk/>

- **Name Changer**

*Name Changer constitue une excellente alternative pour **Mac***

<https://www.macupdate.com/app/mac/21516/namechanger>

12 / RETOUCHER LES FICHIERS IMAGES

CONCEPT-CLÉ

Malgré toutes les précautions prises lors de la prise de vue, il n'est pas rare de devoir faire quelques **retouches en post-production pour parfaire l'image produite ou corriger des anomalies mineures**. Toutes les opérations citées ci-dessous ne sont pas indispensables. C'est à la personne chargée du contrôle – **en s'appuyant sur le cahier des charges ou un document récapitulatif** – de déterminer les éléments à corriger.

Ce document de référence doit comprendre **l'ensemble des paramètres à contrôler** lors de cette étape et les résultats visés pour chaque type de document. Liées au processus, ces informations seront ensuite intégrées aux métadonnées accompagnant les images, ce qui permettra d'en reconstituer l'historique.

La retouche des fichiers images peut comprendre :

- ➔ La **rotation de l'image** pour la placer dans le sens de lecture. Cette opération est particulièrement importante dans le cas d'un document textuel qui doit ensuite être OCRisé. La ligne de texte doit être horizontale, et ce même si la reliure ou le document ne l'est pas tout à fait.
- ➔ Le **détourage**, qui consiste à éliminer les pixels en dehors de la zone de l'objet numérisé.
- ➔ La **correction de la saturation**.
- ➔ La **réduction du bruit**.
- ➔ L'**accentuation de la netteté** ou au contraire l'**adoucissement de l'image**.
- ➔ La **correction du contraste**.
- ➔ La **correction gamma** ou la modification des valeurs de luminance des pixels.

RECOMMANDATIONS

- ➔ Avant de procéder aux retouches, veillez à ce que le **profil ICC du logiciel de traitement d'image soit le même que celui du scanner.**
- ➔ **Travailler sur une copie du fichier numérisé** : il n'est pas possible d'annuler facilement toutes les opérations.

OUTILS

- **Photoshop**

Photoshop est sans doute le plus connu des logiciels de retouche et de traitement d'image. Développé par Adobe, ce logiciel présente de très nombreuses fonctionnalités et est sans doute l'outil le plus complet pour travailler sur une image matricielle.
https://www.adobe.com/be_fr/products/photoshop.html

- **Lightroom**

Lightroom est une alternative intéressante puisque développé spécifiquement pour la retouche de photographies. Il propose également un mode d'édition collaborative, permettant si nécessaire à plusieurs personnes de travailler sur une même image.
https://www.adobe.com/be_fr/products/photoshop-lightroom.html

- **Gimp**

GIMP est un éditeur d'images multiplateforme disponible pour GNU/Linux, OS X, Windows et d'autres systèmes d'exploitation. Moins pointu que les deux programmes précédemment cités, il présente néanmoins l'avantage d'être gratuit et de s'adapter à tous les supports d'exploitation.
<https://www.gimp.org/>

13 / OCÉRISER LES DOCUMENTS TEXTES

CONCEPT-CLÉ

L'océrisation (ou **OCR** pour *Optical Character Recognition* – Reconnaissance Optique de Caractères) est une **technique de traitement de l'image visant à en extraire le texte pour faciliter la lecture de ce dernier et permettre à l'utilisateur d'effectuer aisément des recherches.**

Le principe de l'océrisation se divise en plusieurs étapes qui comprennent :

- ➔ Le **redressement de la page.**
- ➔ La **binarisation du document**, c'est-à-dire sa conversion en noir et blanc.
- ➔ La **segmentation de l'image en zones** de texte, puis en lignes, mots et caractères.
- ➔ Et enfin la **classification de ces zones** par la reconnaissance des caractères puis des mots.

RECOMMANDATIONS

- ➔ Pour contrôler la qualité de l'océrisation, se baser sur un document OCR référentiel, ou vérité-terrain.
- ➔ Pour créer une vérité-terrain, effectuer une océrisation « manuelle » sur un document image :
 - Binariser l'image.
 - Supprimer le bruit automatiquement et corriger éventuellement manuellement les erreurs ou imprécisions de cette correction automatique.

- Supprimer les bordures.
- Créer manuellement ou semi automatiquement les zones de textes.
- Classer les zones selon leur type (image, texte, inscription manuscrite, dessin).
- Définir manuellement l'ordre des lectures entre les blocs de textes du document.
- Définir les lignes de texte et entrer manuellement une phrase représentative, c'est-à-dire l'ensemble le plus hétéroclite de caractères normaux et spéciaux.

➔ Ce document référentiel permet ensuite de juger, sur la base de critères factuels, la faculté de reconnaissance du logiciel des zones de texte, lignes et caractères et notamment :

- Les omissions.
- Les problèmes de classification.
- Les problèmes de segmentation, c'est-à-dire de décomposition de l'image entière en sous-parties – lignes, mots, caractères.

➔ Le contrôle de l'océrisation peut être automatisé à l'aide du logiciel LayoutEval qui, sur la base du référentiel chargé, comparera rapidement les différences de qualité.

OUTILS

- **ABBY FineReader**

ABBY FineReader – Logiciel propriétaire

https://www.abby.com/media/6652/guide_english.pdf

- **Tesseract**

Tesseract – Logiciel open source

<https://ichi.pro/fr/guide-du-debutant-sur-tesseract-ocr-18608940018749>

- **Layout Evaluation**

Layout Evaluation – Logiciel propriétaire

<https://www.primaresearch.org/tools/PerformanceEvaluation>

14 / ENCAPSULER LES MÉTADONNÉES

CONCEPT-CLÉ

Pour être exploitable et diffusable, tout fichier numérique doit comprendre un ensemble de métadonnées décrivant le plus précisément possible le contenu numérique. Ces métadonnées peuvent être externes ou internes mais doivent être rassemblées au sein d'un même document lors de la phase de post-production, en vue de l'optimisation de leur conservation et de leur exploitation.

- ➔ Les métadonnées descriptives comprennent l'ensemble des informations décrivant le contenu numérique -- auteur, date de création du document, sujet, etc. – et sont consignées dans un document distinct.
- ➔ Les métadonnées techniques et administratives reflètent quant à elles toutes les informations liées au processus de numérisation de l'objet : date de capture, modèle de scanner ou de l'appareil photo utilisé, espace colorimétrique du fichier, profil ICC, résolution, taille en pixels, profondeur de codage, etc. Ces métadonnées sont le plus souvent internes et automatiquement générées par le matériel de prise de vue.

RECOMMANDATIONS

Le **standard METS** (*Metadata Encoding and Transmission Standard*) développé par la Digital Library Federation permet d'exprimer en détails toute information liée à la structure interne des documents. Basé sur le schéma XML, ce document permet de **regrouper et de lier tous les éléments liés à la description de l'objet et aux informations concernant les modalités de créations et d'exploitation du contenu numérisé**. Le fichier METS est donc à voir comme un container de métadonnées, pointant vers divers fichiers liés. Il a l'avantage d'être compatible avec d'autres standards de métadonnées comme MARC, EAD ou encore MODS.

Un fichier METS compte sept catégories :

- 1 L'en-tête, décrivant le document METS lui-même.
- 2 Les **métadonnées descriptives** ou le chemin menant vers le fichier contenant ces informations.
- 3 Les **métadonnées administratives**, qui fournissent des informations sur la manière dont les fichiers ont été créés et enregistrés, et sur les droits de propriété intellectuelle concernant l'objet original.
- 4 La **section des fichiers**, qui renseigne sur les versions électroniques de l'objet numérique.
- 5 La **carte de structure**, qui, comme son nom l'indique, dessine une structure hiérarchique pour l'objet numérique, et relie chaque élément de cette structure aux fichiers de contenu et aux métadonnées qui s'y rapportent.
- 6 Les **liens structurels**, permettant de décrire l'existence d'hyperliens entre éléments au sein de la carte de structure.
- 7 Et enfin les **comportements**, qui définissent les modalités d'exécution des différents fichiers reliés au document METS.

OUTILS

- **Documentation générale et guide d'utilisation du standard METS**
http://www.loc.gov/standards/mets/METSOverview.v3_en.html
- **Schéma actuel du standard METS**
<http://www.loc.gov/standards/mets/mets.xsd>
- **Programmes de création et de gestion de fichier METSn**
<http://www.loc.gov/standards/mets/mets-tools.html>
- **Guide complet de la création et du traitement de fichier METS développé par la Bibliothèque nationale de France**
https://www.bnf.fr/sites/default/files/2021-01/ref_num_metadonnees_mets.pdf

15 / DERNIER CONTRÔLE QUALITÉ

RECOMMANDATIONS

À ce stade, le **contrôle qualité** porte sur les images numériques, mais aussi sur l'ensemble des produits de la prestation :

- La **conformité des supports** et des formats de fichiers.
 - Le **nommage** des supports, des dossiers, des fichiers et des images.
 - La **complétude des fichiers images** en regard du nombre d'originaux à reproduire et/ou à l'indexation qui les accompagnent.
 - L'**intégrité de l'indexation** demandée à partir des règles de saisie formulées.
 - L'**adéquation des caractéristiques techniques des fichiers images** aux objectifs attendus, qui doit pouvoir être évalué de manière objective.
- ➔ Le contrôle exhaustif des résultats obtenus n'est envisageable que lorsque le nombre de documents à numériser est limité. Dans la plupart des cas, un **contrôle par échantillonnage est privilégié**. La norme NF ISO 2859-1 fournit une méthodologie pour ce type de contrôle.
- ➔ Le contrôle par échantillonnage se fait toujours **relativement au seuil de qualité minimale**, c'est-à-dire d'un pourcentage maximum d'objets défectueux dans un lot pour que ce lot puisse être considéré comme satisfaisant.
- ➔ La vérification des indicateurs se fait **relativement aux erreurs considérées comme majeures et mineures**.

16 / CHOISIR SON SYSTÈME DE GESTION DE DONNÉES NUMÉRIQUES

CONCEPT-CLÉ

La préservation à long terme de contenus électroniques est l'ensemble des actions, outils et méthodes mis en œuvre pour réunir, identifier, sélectionner, classer, détruire et conserver des contenus électroniques, sur un support sécurisé, dans le but de les exploiter et de les rendre accessibles dans le temps. L'archivage électronique ne se limite pas au stockage et partage de l'information, mais doit également permettre de veiller à l'intégrité et à la confidentialité des contenus numérisés.

Le choix du système qui assurera la gestion et/ou la conservation des documents numérisés est donc fondamental. En premier lieu, il convient de choisir le type de système le plus pertinent selon les documents archivés :

- Si le système est destiné à gérer des documents vivants et modifiables, il convient d'opter pour un système de gestion électronique de documents (GED).
- Au contraire, si les documents sont figés et validés, alors mieux vaut se tourner vers un système d'archivage électronique légal (SAE) .

Le coffre-fort électronique est un élément complémentaire au SAE. Il est d'ailleurs plus commun de parler de composant coffre-fort électronique (CCFN), terme utilisé dans la norme NF Z 42-020.

Attention à ne pas confondre les deux : le composant coffre-fort numérique assure l'intégrité et la traçabilité des objets numériques qui y sont déposés, tandis que le SAE offre des fonctionnalités archivistiques spécifiques : pérennisation des formats, typologie de documents, gestion du cycle de vie, profils d'archivage, etc.

GED

SYSTÈME DE GESTION DE DOCUMENTS

OBJECTIFS :

- Centraliser l'ensemble des documents, quels que soient leur origine et leur format, en un point unique.
- Harmoniser les processus afin de faciliter la gestion des documents au sein d'une institution.
- Gestion quotidienne des documents au sein d'une institution.

FONCTIONNALITÉS :

- Permet la modification et/ou la destruction des documents.
- Permet la gestion des délais de conservation.
- Peut comprendre une structure organisée de stockage, sous le contrôle des utilisateurs.
- Numérisation des documents (si format papier), capture, indexation, classement, stockage, accès et diffusion.

SAE

SYSTÈME D'ARCHIVAGE ÉLECTRONIQUE LÉGAL

OBJECTIFS :

- Assurer l'accès, la confidentialité, l'intégrité, la pérennité, ainsi que la traçabilité de chaque document.
- Garantir un archivage à valeur probante, c'est-à-dire un archivage assurant la conservation d'un document à long terme, dans le respect des contraintes légales et des normes d'archivage.

FONCTIONNALITÉS :

- Empêche la modification et/ou la destruction des documents.
- Contrôle obligatoire des délais de conservation.
- Comprend obligatoirement une structure rigoureuse de classement.

RECOMMANDATIONS

- ➔ Lors du déploiement d'un Système d'archivage électronique, il est conseillé de mettre en place une **politique d'archivage qui définit les contraintes juridiques, opérationnelles, techniques et fonctionnelles, ainsi que les droits d'accès**. Cela permet de renforcer davantage la fonction de sécurité du SAE.
- ➔ SAE et GED sont des **solutions complémentaires** dans la gestion des archives patrimoniales : la GED facilite la gestion des documents courants et le travail collaboratif dans le cadre de chantiers internes de numérisation par exemple, tandis que le SAE offre une solution de conservation pérenne aux contenus numérisés.

- ➔ Privilégier un SAE conforme à la norme NF-Z42-013. Par la complétude du dossier de preuves qu'il fournit (traçabilité des actions réalisées, preuves d'intégrité, existence du document à une date donnée), le SAE garantit l'authenticité et la conformité du document.

OUTILS

- **Norme AFNOR NF Z 42-013**

Publiée par l'AFNOR en 1999 puis révisée en 2001, en 2009 et plus récemment en 2020, la norme NF Z42-013 définit un ensemble d'exigences et de recommandations fonctionnelles, organisationnelles et d'infrastructure pour la conception et l'exploitation d'un système d'archivage électronique (SAE).

Le texte aborde l'organisation autour des fonctions de l'archivage électronique, le versement, la conservation, l'accessibilité, la communication et la gestion des éliminations, l'interopérabilité du SAE, la restitution et la réversibilité.

- **Norme AFNOR NF Z 42-020**

La norme NF Z42-020 se limite à adresser la couche de stockage du coffre-fort numérique et ne traite pas des problématiques telles que l'authentification, la confidentialité, les métadonnées ou l'interfaçage avec les couches métiers.

- **Norme ISO 14641-1 : 2018**

Entièrement basée sur la norme NF Z42-013, la norme internationale ISO 14641-1 définit également les spécifications pour la mise en œuvre d'un Système d'archivage électronique. Son évolution est en revanche indépendante de la norme AFNOR.

- **Norme ISO 14721 : 2012**

Initialement publiée en 2003 et révisée en 2012, la norme ISO 14721 est une transposition de l'OAIS (Open Archive Information System), modèle conceptuel destiné à la gestion, à l'archivage et à la préservation longue de documents numériques.

- **Norme ISO 15489-1 : 2016**

Révisée en 2016, cette norme détaille les concepts et principes pour la gestion des documents à valeur probante des personnes physiques ou morales, privées ou publiques, et définit les caractéristiques qui procurent à un document sa valeur d'autorité, sa valeur probante : authenticité, intégrité, fiabilité et exploitabilité.

ANCTIL, M.-C, LEGENDRE, M., MÜLLER, T., MAILLET, D., BROUSSEAU, K., RENAUD, L., *Recueil de règles de numérisation*, Bibliothèque et Archives nationales du Québec, Bibliothèque nationale de France et Musée canadien de l'histoire, 2014, [En ligne],

<http://collections.banq.qc.ca/bitstream/52327/2426216/1/4671601.pdf>

ACCART, J.-P., *Le Métier de Documentaliste*. Avec la collaboration de Réthy Marie-Pierre. Éditions du Cercle de la Librairie, 2015.

BACHI, V., FRESA, A., PIEROTTI, C., PRANDONI, C., « The digitization age: mass culture is quality culture. challenges for cultural heritage and society », In: Ioannides, M., Magnenat-Thalmann, N., Fink, E., Žarnić, R., Yen, A.-Y., Quak, E. (eds.) EuroMed 2014.

BAROUCH, G., *Révolution des pratiques managériales : le changement par la qualité*, La Plaine Saint-Denis : Afnor Éditions, 2013.

BELLIER, L., « Numérisation, pour une nouvelle approche de la qualité », In: *Bulletin des bibliothèques de France*, Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB), 2014, [En ligne],

<https://bbf.enssib.fr/contributions/numerisation-pour-une-nouvelle-approche-de-la-qualite>

BIBLIOTHÈQUE ET ARCHIVES NATIONALES DU QUÉBEC., *La numérisation des documents administratifs : Méthodes et recommandations : version revue et corrigée*, Montréal : Bibliothèque et Archives nationales du Québec, mai 2012, 35 p., [En ligne],

<https://docplayer.fr/79033783-La-numerisation-des-documents-administratifs.html>

BIBLIOTHÈQUE NATIONALE DE FRANCE., *Techniques et formats de conversion en mode texte*, BnF - Site institutionnel., n.d, [En ligne],

<https://www.bnf.fr/fr/techniques-et-formats-de-conversion-en-mode-texte>

BIZINGRE, J., PAUMIER, J., RIVIÈRE, P., *Les référentiels du système d'information*, Paris : Dunod, 2013.

BOUTET, A. et ROUDAUT, K., « Les enjeux de la numérisation et de l'ouverture d'archives : le point de vue des professionnels » In : *Terminal*, 110 | 2012, p. 27-37

BRUCE, T., HILLMANN, D., *The Continuum of Metadata Quality: Defining, Expressing, Exploiting*, 2004. [En ligne],

https://www.researchgate.net/publication/247818823_The_Continuum_of_Metadata_Quality_Defining_Expressing_Exploiting

BURESI, C., CÉDELLE-JOUBERT, L., *Conduire un projet de numérisation*, Paris : Tec et Doc, 2002, 326 p.

CARBONE, F., OOSTERBEEK, L., COSTA, F., « Extending and adapting the concept of quality management for museums and cultural heritage attractions: A comparative study of southern European cultural heritage managers' perceptions », In: *Tour Manage Perspect*, 2020, 10 p.

CARON, P., DEFIOLE, R., et Lay, M.-H., *L'enjeu des métadonnées dans les corpus textuels : Un défi pour les sciences humaines*, Presses universitaires de Rennes, 2019.

CHARDONNENS, A., *La gestion des données d'autorité archivistiques dans le cadre du Web de données*, [Thèse de doctorat non publiée], Université Libre de Bruxelles, 2021, 420 p.

<https://difusion.ulb.ac.be/vufind/Record/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/315804/Holdings>

CHAUMIER, J., *Document et numérisation. Enjeux techniques, économiques, culturels et sociaux*, Paris : AdBS éditions, 2006, 119 p.

CHIRON, G., DOUCET, A., COUSTATY, M., VISANI, M., and MOREUX, J.-P., « Impact of OCR Errors on the Use of Digital Libraries: Towards a Better Access to Information », In: *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2017, pp. 1-4.

CLAERR, T., WESTEEL, I., MELOT, M., *Manuel de la numérisation*, Paris : Éd. du Cercle de la, 2011. 317 p.

CLAERR, T., WESTEEL, I., *Numériser et mettre en ligne*, Villeurbanne : Presses de l'enssib, 2010. [En ligne],
<http://books.openedition.org/pressesenssib/414>.

CLAERR, T., WESTEEL, I., ORY-LAVOLÉE, B., GEORGES, N., *Manuel de constitution de bibliothèques numériques*, Paris : Éd. du Cercle de la librairie, 2013. 407 p.

CHEVRY, E., *Stratégies numériques*, Paris : Hermès Science publications-Lavoisier, 2011.

COMMISSION NATIONALE INFORMATIQUE ET LIBERTÉS. *Comment permettre à l'homme de garder la main? Les enjeux éthiques des algorithmes et de l'intelligence artificielle. Synthèse du débat public animé par la CNIL dans le cadre de la mission de réflexion éthique confiée par la Loi pour une république numérique*, décembre 2017, [En ligne],
https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf

CONSEIL CANADIEN DES ARCHIVES. *Manuel de conservation des documents d'archive*, Chapitre 3 – Environnement, 2003, [En ligne],
http://www.cdncouncilarchives.ca/RBch3_fr.pdf

DANGERFIELD, M.-C. et KALSHOVEN, L., *Report and Recommendations from the Task Force on Metadata Quality*, 54 p., [En ligne],
https://pro.europeana.eu/files/Europeana_Professional/Publications/Metadata%20Quality%20Report.pdf

DIRECTION DES ARCHIVES DE FRANCE. MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION - COMITÉ DE PILOTAGE NUMÉRISATION - MISSION DE LA RECHERCHE ET DE LA TECHNOLOGIE (DDAI), *Écrire un cahier des charges de numérisation du patrimoine. Guide technique. Documents reliés- manuscrits- plans- dessins- photographies- microformes*, Paris : Archives de France, 2008, 62p. [En ligne],
<http://www.archivesdefrance.culture.gouv.fr/static/4132>

DIRECTION INTERMINISTÉRIELLE DU NUMÉRIQUE ET DU SYSTÈME D'INFORMATION ET DE COMMUNICATION DE L'ÉTAT, *Référentiel général d'interopérabilité*, 2015, version 2, [En ligne],
https://www.numerique.gouv.fr/uploads/Referentiel_General_Interoperabilite_V2.pdf

DUFRÊNE, B., IHADJADENE, M., DENIS, B., BRUCKMANN, D., BARBIER, B., et RACINE, B., *Numérisation du patrimoine : quelles médiations ? quels accès ? quelles cultures ?*, Paris : Hermann, 2013. 311 p.

D'HOORE, M., *La numérisation du patrimoine écrit : du virtuel à la réalité*, Bruxelles : Archives et bibliothèques de Belgique, 2011. 280 p.

EAD, *EAD en bibliothèque, guide des bonnes pratiques, Origine - origination*, 2020, [En ligne],
<https://www.ead-bibliotheque.fr/guide/donnees-du-did/origination/>

EZRATI, J.-J., *Éclairage d'exposition, musée et autres espaces*, Paris : Eyrolles, 2015.

FEHAU E., LE DANTEC N., *Vade-mecum de la conservation préventive*, C2RMF, version du 18 novembre 2013, [En ligne],
https://c2rmf.fr/sites/c2rmf.fr/files/vademecum_cc.pdf

FEYLER, F., « De la compatibilité à l'interopérabilité en matière de repérage d'information pertinente. La problématique et l'exemple d'OTAREN », In : *Documentaliste-Sciences de l'Information*, vol. 44, no. 1, 2007, pp. 84-92.

FLECKENSTEIN, M., FELLOWS, L., « Data Quality », In : *Modern Data Strategy*, Cham: Springer, 2018.

FONDS INDÉPENDANT DE PRODUCTION. *Êtes-vous repérables? Guide pratique pour documenter vos contenus*, novembre 2017, [En ligne],
<http://ipf.ca/FIP/ressources/etes-vous-reperables/#1512498362849-3b320f4f-a27f>

FONDS DES MÉDIAS DU CANADA, *Découvrabilité : vers un cadre de référence commun*, 2016. [En ligne],
<https://trends.cmf-fmc.ca/fr/rapports-de-recherche/decouvrabilite-vers-un-cadre-de-reference-commun>

FORCE 11, *The Fair Data Principles*, n.d. [En ligne],
<https://www.force11.org/group/fairgroup/fairprinciples>

GIESEN, E., *Démarche qualité et norme ISO 9001 : Une culture managériale appliquée à la recherche*, Marseille : IRD Éditions, 2018, [En ligne],
<http://books.openedition.org/irdeditions/20609>

GILLET, F., HUNGENAERT, J., HODZA, M., et al., *Identifying Needs to Modernize Access to Digital Data in Libraries and Archives (MADDLAIN) : final report*, Brussels : Belgian scientific Policy, 2018, [En ligne],
http://www.belspo.be/belspo/brain-be/projects/FinalReports/MADDLAIN_%20final%20report.pdf

GUEGUEN, G., MARQUES DA FONSECA, V.-M., D., PITTI, V. et SIBILLE, C., *Vers un modèle conceptuel international pour la description archivistique. Toward an International Conceptual Model for Archival Description: A Preliminary Report from the International Council on Archives' Experts Group on Archival Description*, 2013, [En ligne],
https://www.ica.org/sites/default/files/EGAD_French.pdf

INTERNATIONAL CENTRE FOR THE STUDY OF THE PRESERVATION AND RESTORATION OF CULTURAL PROPERTY (ICCROM), *A Guide to Risk Management of Cultural Heritage*, 2016, [En ligne],
<https://www.iccrom.org/publication/guide-risk-management>

INTERNATIONAL CENTRE FOR THE STUDY OF THE PRESERVATION AND RESTORATION OF CULTURAL PROPERTY (ICCROM), *The ABC Method. A risk management approach to the preservation of cultural heritage*, 2016, [En ligne],
<https://www.iccrom.org/publication/abc-method-risk-management-approach-preservation-cultural-heritagf>

INSTITUT DE CONSERVATION DU CANADA (ICC), *Plan de préservation des collections patrimoniales : Stratégies pour éviter ou limiter les dommages*, n.d., [En ligne],
<https://www.canada.ca/fr/institut-conservation/services/conservation-preventive/plan-preservation-collections-patrimoniales.html>

INSTITUT ROYAL DU PATRIMOINE ARTISTIQUE (IRPA), *Cellule de conservation préventive*,
<https://www.kikirpa.be/fr/conservation-restauration/cellule-conservation-pr%C3%A9ventive>

KULESZ, O., *Culture, machines et plateformes : L'intelligence artificielle et son impact sur la diversité des expressions culturelles*, Comité intergouvernemental pour la protection et la promotion de la diversité des expressions culturelles, Douzième session, Paris, siège de l'UNESCO, 11-14 décembre 2018, DCE/18/12.IGC/INF.4, 13 novembre 2018, [En ligne],
https://en.unesco.org/creativity/sites/creativity/files/12igc_inf4_fr.pdf

LU, T., *Homogeneity models for image processing in the cultural heritage sector*, [Thèse de Doctorat non publiée], Vrije Universiteit Brussel, 2020, 245 p.

MANDERLKERN, D., *Diffusion des données publiques et révolution numérique*, Rapport rédigé à la suite des travaux de l'Atelier présidé par D. Mandelkern au Commissariat au Plan, Paris : La Documentation Française, 1999, 124 p.

MEMON, J., SAMI, M., KHAN R.-A., UDDIN M., « Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR) », In : *IEEE Access*, vol. 8, blz. 142642-142668, 2020.

MICKELVEY, F., HUNT, R., *Responsabilité algorithmique et découverte de contenus numériques, document de réflexion, Rencontre internationale sur la diversité des contenus à l'ère numérique*, Patrimoine canadien, février 2019, [En ligne],
<https://fr.readkong.com/page/responsabilit-algorithmique-et-d-couverte-de-contenus-2696781>

MINISTÈRE DE LA CULTURE ET DE LA COMMUNICATION, *Feuille de route stratégique : Métadonnées culturelles et transition Web 3.0*, janvier 2014, 38 p. [En ligne],
<https://www.enssib.fr/bibliotheque-numerique/documents/64776-feuille-de-route-strategique-metadonnees-culturelles-et-transition-web-3-0.pdf>

NAPOLI, P., *Diversité de contenus à l'ère numérique : découvrabilité de contenu diversifié aux échelons local, régional et national, document de réflexion, Rencontre internationale sur la diversité des contenus à l'ère numérique*, Patrimoine canadien. Février 2019, [En ligne],
<https://culturenumeriqc.qcnum.com/wp-content/uploads/2019/03/Napoli-De%CC%81couvrabilite%CC%81-de-contenu-diversifie%CC%81-aux-e%CC%81chelons-local-re%CC%81gional-et-national.pdf>

PLAN D'EXPLOITATION ET DE PRÉSERVATION DES PATRIMOINES (PEP'S). *Normes et lignes directrices techniques et organisationnelles pour les initiatives de numérisation des patrimoines culturels soutenues par la Communauté française*, ministère de la communauté française Bruxelles : Secrétariat général – Préservation et exploitation des patrimoines, 2009, 61 p. [En ligne],
http://www.peps.cfwb.be/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/sites/numpat/upload/numpat_super_editor/numpat_editor/documents/CFWB/Normes.pdf&hash=a2d0e740e46e34b8c3475c604c3e96edbb8a8712

OBSERVATOIRE DE LA CULTURE ET DES COMMUNICATIONS DU QUÉBEC.

État des lieux sur les métadonnées relatives aux contenus culturels, Québec :

Institut de la statistique du Québec, Observatoire de la culture et des communications du Québec, 2017, 118 p., [En ligne],

www.stat.gouv.qc.ca/observatoire

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Information et documentation : Partie 1 : Principes directeurs*, Genève : ISO, ISO 15489-1 : 2001, 2001.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Information et documentation : Partie 2 : Guide pratique*. Genève : ISO, ISO 15489-2 : 2001, 2001.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Photographie – Illuminants sensitométriques – Spécifications pour la lumière du jour, la lumière artificielle et la tireuse*. Genève : ISO, ISO 7589 :2002, 2002.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Photographie – Scanners électroniques pour images photographiques – Mesurages d'intervalles dynamiques*. Genève : ISO, ISO 21550: 2004, 2004.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Gestion de documents : Format de fichier des documents électroniques pour une conservation à long terme : Partie 1 : Utilisation du PDF 1.4 (PDF/A-1)*. Genève : ISO, ISO 19005-1, 2005.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Photographie – Appareils de prises de vue électroniques – Méthodes de mesure des fonctions de conversion opto-électroniques*. Genève : ISO, ISO 14524: 2009, 2009.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*. Genève : ISO, ISO 14721: 2012, 2012.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Systèmes de management de la qualité – Principes essentiels et vocabulaire*. Genève : ISO, ISO 9001 :2015, 2015.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Information and documentation – Records management – Part 1: Concepts and principles*. Genève : ISO, ISO 15489:2016, 2016.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Systèmes de management de la qualité – Exigences*. Genève : ISO, ISO 9004 :2018, 2018.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Quality management – Quality of an organization – Guidance to achieve sustained success.* Genève : ISO, ISO 9004 :2018, 2018.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Electronic document management – Design and operation of an information system for the preservation of electronic documents – Specifications.* Genève : ISO, ISO 114641 :2018, 2018.

ORGANISATION INTERNATIONALE DE NORMALISATION (ISO). *Information et documentation – L'ensemble des éléments de métadonnées Dublin Core – Partie 2.* Genève : ISO, ISO 15836-2 :2019, 2019.

PLAMONDON, J., *Bien documenter pour favoriser la découverte en ligne. Travailler avec des données,* 2019, [En ligne],
<https://espaceschoregraphiques2.com/wp-content/uploads/2019/12/Guide-me%CC%81adonne%CC%81es-FJPP.pdf>

PLAMONDON, J., *Données ouvertes et liées : le Web comme base de données,* mars 2018, [En ligne],
<https://joseplamondon.com/donnees-ouvertes-liees-web-base-de-donnees/>

PLAMONDON, J., *Produire des données : entre outils de marketing et bases de connaissances,* 21 août 2019, [En ligne],
<https://joseplamondon.com/produire-donnees-culture-marketing-ou-bases-connaissances>

RIEGER, T., *Federal Agencies Digital Guidelines Initiative. Technical Guidelines for Digitizing Cultural Heritage Material. Creation of Raster Image File,* 2016, 99 p., [En ligne],
http://www.digitizationguidelines.gov/guidelines/FADGI%20Federal%20Agencies%20Digital%20Guidelines%20Initiative-2016%20Final_rev1.pdf

SALAÜN. J.-M., *Bibliothèques numériques et Google-Print, Regards sur l'actualité : mensuel de la vie publique en France,* La Documentation Française, 2005, [En ligne],
https://archivesic.ccsd.cnrs.fr/sic_00001576/document

SIBILLE, C., « Avantages et inconvénients d'un encodage en EAD », *In: La Gazette des archives*, n°220, 2010-4, Les instruments de recherche : évolutions, publics et stratégies, pp. 179-194.

SOYEZ, S., *La numérisation en marche : les étapes de la dématérialisation des processus de travail,* Bruxelles : Archives générales du Royaume et Archives de l'État dans les Provinces, n.d., [En ligne],
http://www.arch.be/docs/brochures/la_numerisation_en_marche.pdf

TCHÉHOUALI, D., « Note d'orientation pour une étude sur les enjeux, défis et opportunités de la découvrabilité en ligne des contenus culturels francophones », *In: Dispositif d'observation des dynamiques culturelles et linguistiques*, Comité d'orientation des Dynamiques culturelles et linguistiques, avril 2018, [En ligne], https://www.francophonie.org/sites/default/files/2020-01/perspectives_de_recherche_web-2.pdf

TESSIER, M., FERRY, L., RACINE, B., JEANNENEY, J.-N., *La révolution du livre numérique : état des lieux, débats, enjeux*, Paris : Odile Jacob, 2011. 208 p.

THIAULT, F., « Mutations des métiers de l'information-documentation : vers l'émergence d'une culture de l'information numérique », *In: Les Cahiers de la SFSIC*, Société française des sciences de l'information et de la communication, 2012, p. 59-62.

VAN DORMOLEN, H., *Metamorphoze Preservation Imaging Guidelines. Image Quality*, 2012, version 1.0, [En ligne], http://www.imagingetc.com/images/Resources/Images/PDFs_DownloadFiles/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf

VILLANUEVA, E., et SHIRI, A., « Methodological Diversity in the Evaluation of Cultural Heritage Digital Libraries and Archives: An Analysis of Frameworks and Methods / Diversité méthodologique dans l'évaluation des bibliothèques et les archives numériques du patrimoine culturel : Une analyse des cadres et des méthodes », *In: Canadian Journal of Information and Library Science*, vol. 43 no. 3, 2021, p. 316-342.

