



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Office fédéral de la protection de la population OFPP

> 03.2008

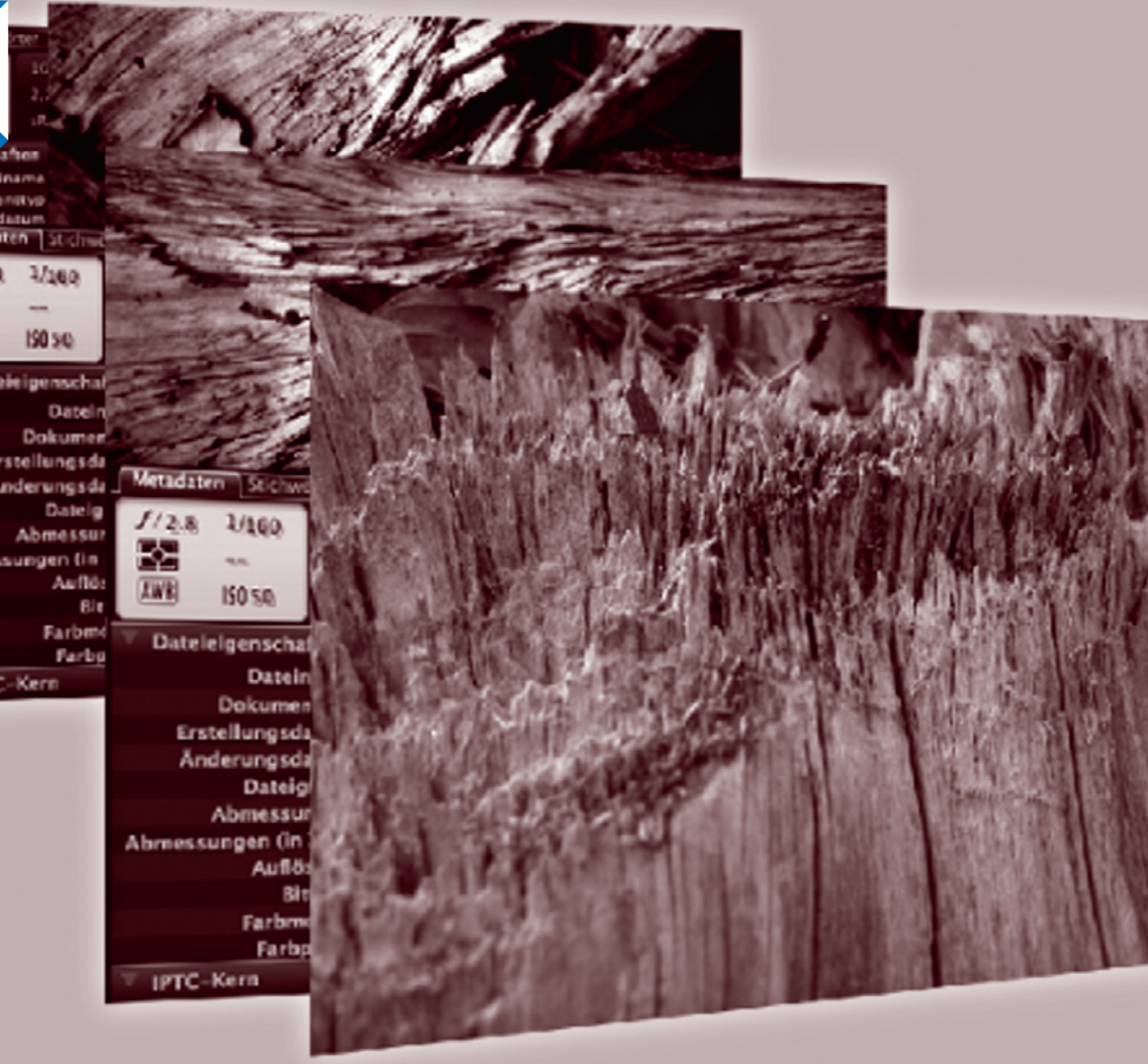
> THEMA: DIRECTIVES DE LA PBC CONCERNANT LES MÉTADONNÉES DES
IMAGES FIXES NUMÉRIQUES

> AUTEURS: SERGIO GREGORIO, ANJA-ELENA STEPANOVIC



GUIDELINES

DIRECTIVES DE LA PBC CONCERNANT LES MÉTA- DONNÉES DES IMAGES FIXES NUMÉRIQUES



COUVERTURE

Illustration de formats d'images courants dans les banques de données.
Photo: Elias Kreyenbühl, Imaging and Media Lab

DIRECTIVES DE LA PBC

CONCERNANT LES MÉTA-

DONNÉES DES IMAGES

FIXES NUMÉRIQUES

AUTEURS :

SERGIO GREGORIO & ANJA-ELENA STEPANOVIC

Etat: 28.02.2008

RESPONSABLES DU PROJET ET DE LA GESTION DU PROJET:

RINO BÜCHEL, RETO SUTER

Mandant:

© Office fédéral de la protection de la population OFPP, Protection des biens culturels PBC, Berne, 2008

Rédaction:

Reto Suter, Hans Schüpbach

Traduction:

Muriel Baudat

Mise en page:

Centre des médias électroniques CME

AVANT-PROPOS

Rino Büchel

Chef de la section Protection des biens culturels PBC dans l'Office fédéral de la protection de la population OFPP

Prof Dr Rudolf Gschwind

Imaging and Media Lab, Bâle

En règle générale, la définition des métadonnées consiste en une paraphrase simplifiée: «données sur des données». Toutefois, sans rapport à la pratique, cette brève description ne dit pas grand-chose. Dans le contexte numérique, les métadonnées décrivent des objets numériques (fichiers textes, images ou sons), rendent ceux-ci plus faciles à retrouver et appuient la transmission des informations. Bien que des documentations détaillées concernant divers aspects soient disponibles sur Internet, le thème des métadonnées reste, même pour l'utilisateur intéressé, abstrait, diffus et parfois difficile à concevoir. Un exemple tiré du monde physique devrait permettre de mieux illustrer cette notion.

Scénario: après une inondation, l'eau est pompée dans les caves et les locaux sont nettoyés et remis en état. Les provisions qui ne peuvent plus être sauvées sont éliminées. Mais que faire du tas de boîtes de conserve en vrac, dont l'eau a retiré les étiquettes? Leur contenu a survécu à la catastrophe sans dommage, mais de l'extérieur, sans ouvrir les boîtes, il est impossible de savoir ce qu'elles renferment. Comment, par exemple, distinguer des raviolis d'une boîte d'aliments pour chien?

Il en va de même dans le monde numérique. Un document, dans notre cas une image numérique qui n'a aucune description – et ne contient donc aucune métadonnée – doit d'abord être ouvert pour que l'on puisse identifier de quoi il s'agit. Cependant, il y a aussi des boîtes dont la description du contenu est imprimée sur le contenant métallique (voir



L'exemple dans les fig. 1 et 2 montre la nécessité des métadonnées pour l'identification des objets.



boîte de conserve à droite de l'image) et reste donc encore visible après les dégâts d'eau. Ces boîtes sont plus chères, l'impression constituant une plus-value. Par analogie, il existe dans le monde numérique des métadonnées qui sont intégrées à l'objet, impliquant ainsi un investissement supplémentaire et une augmentation des coûts de production. Dans notre exemple, les étiquettes en papier correspondent à des métadonnées enregistrées de manière externe (la plupart du temps dans un fichier indépendant ou dans une banque de données), qui seront très probablement perdues si elles sont séparées de l'objet. La comparaison s'arrête là, ne pouvant pas prendre en compte tous les détails. Ce qui est important, du point de vue de l'archivage à long terme surtout, c'est que les objets numériques soient pourvus de métadonnées, non seulement pour qu'ils soient faciles à retrouver, mais aussi pour qu'ils restent interprétables à l'avenir.

Tout comme dans le monde physique, les métadonnées (l'inscription, l'étiquette) du monde numérique appartiennent aux objets et devraient, dans la mesure du possible, être directement intégrées au fichier auquel elles correspondent.

Dans le domaine des images numériques, de nombreuses possibilités sont encore ouvertes et le problème des métadonnées n'a pas encore trouvé de solution satisfaisante. Cet état de fait ne doit toutefois pas inciter à négliger l'enregistrement minutieux des images numériques sur un support de données et à attendre «des temps meilleurs». Si l'on procédait de la sorte, le temps qui passe verrait augmenter de plus en plus la probabilité et le risque de perdre des données. Seule une exploitation active des images numériques garantit aujourd'hui la pérennité des biens culturels numériques.

Les présentes directives sont une introduction au thème des métadonnées d'images numériques fixes; elles expliquent un procédé de définition des métadonnées qui a fait ses preuves et présentent les étapes que devraient suivre les images numériques en vue de leur archivage à long terme. Nous attirons votre attention sur la liste de contrôle du chapitre 10.1, qui peut aider les responsables des institutions concernées à s'orienter.

De nombreuses personnes, citées ci-dessous par ordre alphabétique, ont apporté des commentaires à l'ébauche de ces directives – leurs remarques ont été incluses lors de l'élaboration de la version finale. Nous remercions sincèrement Kurt Deggeller (directeur de Memoriav), Urspeter Schelbert (archiviste d'Etat, canton de Zoug), Bernard A. Schüle (responsable du Centre des objets des Musées nationaux suisses), Daniel Stadlin, responsable PBC du canton de Zoug, et Tobias Wildi (Docuteam GmbH) pour le regard critique qu'ils ont porté à ce document.

Ces directives ont pour but d'assister les institutions culturelles de petite et de moyenne taille (archives, bibliothèques, musées et institutions parentes) dans l'exploitation et l'archivage de leurs images numériques. La Protection des biens culturels de l'Office fédéral de la protection de la population (PBC/OFPP) apporte ainsi, en collaboration avec l'Imaging & Media Lab de l'Université de Bâle, une contribution fondamentale à la lutte contre l'oubli numérique.

TABLE DES MATIÈRES

1	Introduction	8
2	Formats des fichiers images	9
2.1	TIFF	9
2.2	JPEG	10
2.3	JPEG2000	11
2.4	Conclusions	11
3	Métadonnées	12
3.1	Classification	13
3.2	Conclusions	14
4	Standards de métadonnées	15
4.1	Conclusions relatives aux métadonnées d'images	17
5	Exemple d'application	18
6	Champs principaux	23
7	Archivage numérique	24
7.1	Processus d'archivage	25
7.1.1.	Préparation	25
7.1.2.	Traitement	25
7.1.3.	Copie	26
7.1.4.	Administration	27
7.2	Conclusions	27
8	Conclusions finales et perspectives	28
9	Bibliographie	30
10	Annexe	32
10.1	Liste de contrôle	32
10.2	Champs principaux RLG (Summary of Core fields)	34
10.3	Vue d'ensemble: normes et schémas de métadonnées	35
10.3.1.1–10.3.1.21	Normes et schémas en détail	35
10.4	Vocabulaires et thésaurus contrôlés	41
10.4.1.1–10.4.1.11	Vocabulaires et thésaurus en détail	41
11	Liste des abréviations	44

1 INTRODUCTION

¹ Dans la suite du texte, le terme *images fixes numériques* (*digital still images*) sera remplacé par sa version abrégée, *images numériques*.

Le nombre croissant d'objets et le flot en constante augmentation de documents disponibles exclusivement sous forme numérique rendent difficiles l'obtention d'une vue d'ensemble et la recherche ciblée d'informations. Souvent, les résultats des recherches sont incomplets en dépit de l'utilisation des moyens les plus modernes, car sans description, il est difficile de trouver un objet numérique. Cette réalité touche en particulier les images fixes numériques¹, dont le contenu ne peut pas être retrouvé ultérieurement au moyen d'une simple recherche textuelle. C'est la raison pour laquelle les descriptions d'images sont d'une importance fondamentale pour les images numériques. Des métadonnées descriptives représentent également une condition sine qua non pour la conservation du contexte des objets archivés et permettront à l'avenir, en tant qu'élément constitutif d'un objet numérique indissociable, non seulement de retrouver les données, mais aussi d'en comprendre le contexte historique.

Alors qu'une description pertinente des objets peut améliorer les résultats de la recherche d'informations, on accorde en général peu d'attention à l'aspect de l'archivage numérique. L'utilisation relativement simple des technologies de l'information induit en erreur sur la conservation à long terme des objets numériques. Tant qu'elles sont accessibles et interprétables, on admet en effet que les données numériques peuvent être conservées à long terme. Sans une stratégie d'archivage numérique orientée sur une longue durée, la perte des données et des informations qui y sont rattachées est inéluctable. C'est pourquoi cet aspect doit forcément être pris en compte dans le traitement des données numériques.

Les présentes directives constituent une introduction au thème des métadonnées des images numériques et montrent comment munir les fichiers images d'informations en vue de leur conservation à long terme. Les différentes étapes sont illustrées au moyen d'un processus d'archivage. En introduction, nous présenterons les principaux formats de fichiers images actuels et expliquerons les différents types de métadonnées. Nous donnerons ensuite un aperçu des normes actuelles en matière de métadonnées, parmi lesquelles il n'y a pas pour l'instant de norme de jure pour les images numériques. Les procédés analogiques tels que l'enregistrement sur microfilm ne seront pas mentionnés et nous n'aborderons que de manière marginale les aspects de la gestion des écrits. Les explications ci-dessous sont avant tout destinées à fournir un appui aux petites et moyennes institutions culturelles pour ce qui relève de l'administration, de la gestion et de l'archivage de leurs images.

2 FORMATS DES FICHIERS IMAGES

² Toutes les abréviations utilisées dans le présent document sont expliquées dans la liste des abréviations.

³ Pour obtenir un aperçu très détaillé des formats de fichiers images actuels et obsolètes, consultez la page: http://www.lemkesoft.de/xd/public/content/index._cGkPTY5_.html.

⁴ Un standard propriétaire appartient à une société, qui ne révèle pas ses spécifications du tout, ou seulement en partie (p. ex. le format actuel de document Word de Microsoft, version Word 2003).

⁵ L'extension usuelle est «PSD». Le format Photoshop supporte, entre autres, des plates-formes et des profils de couleurs.

⁶ Le format TIFF est constitué de «champs» descriptifs (tag ou étiquette), qui contiennent une valeur, p. ex. la taille de l'image, sa résolution, etc.

⁷ Points visuels (pixels) organisés en lignes et en colonnes.

⁸ Le format TIFF a été créé en 1992 par l'entreprise Aldus Corporation, qui, avec Aldus Pagemaker, compte parmi les premiers fournisseurs de logiciels de composition de pages. En 1994, elle a fusionné avec Adobe.

⁹ Les spécifications TIFF peuvent être téléchargées sur le site Web d'Adobe, à l'adresse suivante: <http://partners.adobe.com/asn/developer/PDFS/TN/TIFF6.pdf>. Le format TIFF peut être élargi au moyen de tags personnalisés (private tags).

Un format d'image est une suite précise de bits (0 et 1) appelée flux binaire. Cette suite est définie avec précision afin que l'ordinateur puisse identifier s'il s'agit d'un fichier texte, image ou son et le traiter comme tel. Il existe des spécifications pour certains formats de fichiers, qui décrivent leurs caractéristiques.

Pour les images numériques, il existe de nombreux formats de données. En principe, seuls les formats JPEG² et TIFF sortent du lot pour l'utilisation et la conservation à long terme³. Pour l'archivage aussi bien que pour l'utilisation, il est important qu'il s'agisse de formats ouverts, reposant sur un standard ouvert. Un standard ouvert, comparé à un standard propriétaire⁴, est documenté et compréhensible dans ses moindres détails.

Aujourd'hui et selon les pratiques actuelles, choisir d'utiliser des formats JPEG et TIFF pour la conservation à long terme, c'est choisir la sécurité. Naturellement, d'autres formats, comme ceux d'Adobe Photoshop⁵ sont également autorisés pour le traitement des données. Toutefois, dès que celui-ci est terminé, il est recommandé d'en convertir le résultat en un format adapté à l'archivage; avec les cycles d'innovations de plus en plus rapides, les formats propriétaires des fournisseurs de logiciels commerciaux changent de plus en plus vite et sont par conséquent de plus en plus rapidement illisibles. Il faudrait par exemple, après avoir terminé toutes les étapes nécessaires de traitement de l'image, créer une ou plusieurs versions JPEG destinées à l'utilisation, ainsi qu'une version TIFF destinée à l'archivage.

2.1 TIFF

TIFF est l'acronyme de Tagged Image File Format et est un format qui se fonde sur des descriptions ou des tags⁶ et a été conçu pour le traitement d'images binaires⁷ et de graphiques.

Le format TIFF – dont l'extension de fichier usuelle est .tif – existe depuis 1992⁸ et a continué d'être développé depuis. Les nouvelles exigences ayant été accompagnées de nouvelles balises, le format TIFF englobe aujourd'hui plusieurs standards⁹.

Les avantages du format TIFF sont sa large diffusion, une certaine solidité (relativement peu sujet à erreur relativement faible) et le fait qu'il s'agit d'un format dit non compressé. Concrètement, pour l'utilisateur, cela signifie par exemple que les enregistrements successifs ne provoquent aucune perte d'information; un fichier TIFF est modifiable à volonté, sans que des informations de l'image ne soient perdues, ce qui ne dit toutefois rien du contenu visuel de l'image.

Les inconvénients de ce format sont sa complexité et surtout la taille des fichiers, qui ont pour conséquence une charge financière importante si les données sont nombreuses. Pour l'archivage durable des images numériques, TIFF reste le format de fichier qui s'impose, bien que la compression des données soit de plus en plus en point de mire dans les discussions sur l'archivage à long terme des données numériques, les considérations économiques ne pouvant être ignorées.

¹⁰ JPEG a été normalisé en 1990 et a ensuite connu un développement croissant grâce à l'appui et à l'utilisation des fournisseurs de logiciels. Ses spécifications sont disponibles à l'adresse suivante: <http://www.w3.org/Graphics/JPEG/jfif3.pdf>.

¹¹ Le terme IPTC est défini au chapitre 3.

¹² Si les images d'archives au format JPEG doivent être retravaillées, il est recommandé de les convertir d'abord au format TIFF et, au terme du travail, de recréer des versions JPEG si nécessaire. On s'assure ainsi de ne pas perdre de données lors de la compression avec pertes JPEG.

2.2 JPEG

JPEG est l'acronyme de «Joint Photographic Experts Group», un format qui a connu une très large diffusion ces derniers temps, en particulier grâce à l'explosion du nombre d'utilisateurs d'appareils photo numériques¹⁰. JPEG – dont l'extension de fichier usuelle est .jpg – est un format qui compresse les données en induisant une perte, et qui, comparé au format TIFF, allie une taille de fichier réduite à une bonne qualité visuelle. Pour l'utilisateur, cela signifie que chaque nouvel enregistrement d'un fichier JPEG provoque une perte d'information. Celle-ci peut être constatée par la diminution de la taille du fichier et l'augmentation du nombre d'artéfacts de l'image, sous la forme de mosaïques clairement visibles et gênantes. À l'origine, il s'agissait de créer un format d'image de bonne qualité visuelle et relativement mince pour faciliter la transmission de données.

Les avantages du format JPEG sont sa large diffusion et la lisibilité à plus long terme qui en résulte implicitement, ainsi que la saisie simple des métadonnées IPTC¹¹. Parmi ses inconvénients, on citera la perte d'informations due à la compression et son caractère moins robuste que le format TIFF: à l'intérieur d'un fichier JPEG, les erreurs sont visibles immédiatement.

En tant que format induisant des pertes d'informations, le format JPEG n'est en principe pas recommandé pour un archivage durable de données numériques. Toutefois, pour des raisons économiques, la compression de données est envisagée de plus en plus souvent même pour l'archivage numérique à long terme¹².

¹³ Outre la complexité intrinsèque du format JPEG2000, il se peut que le fait que le moteur de recherche le plus célèbre ne supporte pas ce format soit également responsable de sa faible diffusion. Une liste des paquets de logiciels supportant le format JPEG2000 est disponible sur: http://de.wikipedia.org/wiki/JPEG_2000.

¹⁴ L'explication de ces aspects principalement techniques dépasserait le cadre de la présente publication. Wikipédia propose une bonne vue d'ensemble et des explications de qualité à ce sujet. Cf. http://fr.wikipedia.org/wiki/Transform%C3%A9e_en_ondelette_discr%C3%A8te.

2.3 JPEG2000

Depuis quelques années, le nouveau format de fichier image JPEG2000 - une version perfectionnée de JPEG - est présenté comme le futur format des fichiers d'archives et la solution de remplacement du format TIFF. Un examen détaillé des propriétés de ce nouveau format nous montre que les attentes et les espoirs qu'il suscite sont absolument justifiés. Ce format n'est pas encore très répandu car peu de logiciels le supportent totalement pour l'instant¹³.

JPEG2000 est un format de fichiers images entièrement nouveau, qui n'a plus rien de comparable avec JPEG et TIFF. Il se fonde sur ce que l'on appelle la transformée en ondelettes discrètes, qui construit les images progressivement et peut utiliser indifféremment des méthodes de compression avec pertes ou sans perte d'information¹⁴. Les avantages de JPEG2000 résident dans la plus faible perceptibilité des artéfacts (mosaïques gênantes) au niveau visuel et dans la grande robustesse (occurrence des erreurs) au niveau du format, ce qui est important pour la durabilité de l'archivage.

Si la création directe d'images JPEG2000 ne devrait pas poser problème, il faudra par contre veiller, en cas de migration future, à ce qu'il n'y ait aucune perte de données ou de qualité et à ce que la migration soit réversible.

2.4 CONCLUSIONS

Avec l'archivage d'images au format TIFF, on continue de miser sur la sécurité. A l'heure actuelle, il y a cependant des images numériques qui sont transmises directement à partir d'appareils photo numériques, presque exclusivement au format JPEG. L'investissement nécessaire pour convertir celles-ci au format TIFF avant l'archivage peut être très élevé, en particulier lors de grandes quantités d'images. Ainsi, même si les principes prescrivent un format adapté aux archives tel que TIFF, dans la pratique l'archivage de fichiers JPEG peut s'avérer tout à fait pertinent si les travaux de préparation sont réalisés correctement et qu'une migration ultérieure de format, p. ex. vers JPEG2000, est entreprise en temps opportun et en prenant toutes les mesures nécessaires.

Une migration ou une conversion de format est requise lorsque le format devient obsolète, c.-à-d. lorsque de nouveaux logiciels ne supportent plus ce format, ou que le nouveau format présente des avantages évidents par rapport à l'ancien. D'autres points essentiels, en particulier dans le domaine des images numériques, sont présentés dans Rosenthaler (2007).

3 MÉTADONNÉES

¹⁵ «Metadata is structured information that describes, explains, locates, or otherwise makes it easier to manage an information resource».

¹⁶ «A resource may be a publication such as a report or journal, a web page about a particular topic or service, or a digital object such as an image»: <http://www.nla.gov.au/guidelines/metaguide.html>.

¹⁷ «[...] any formal scheme of resource description, applying to any type of object, digital or non-digital»

¹⁸ D'autres aspects établissant plus précisément la distinction entre les données primaires (objets de données) et les données secondaires (métadonnées) sont présentés dans Margulies (2007).

¹⁹ La norme ISO 15836:2003 ne définit par exemple que les éléments du Dublin Core (Dublin core element set), qui peuvent être utilisés.

²⁰ La norme ISO 15489, Information et documentation, «Records management», se compose de deux parties, une partie générale, ISO 15489-1 (utilisation, champ de réglementation, principes, prescriptions quant aux objectifs, responsabilités, exigences, systèmes de gestion des documents d'archive, procédures et gestion, examen, formation et perfectionnement) et une partie spécifique, Guide pratique ISO 15489-2 (champ d'application, principes et compétences, stratégie, développement, mise en œuvre, surveillance et formation). La norme ISO 23081 décrit les métadonnées pour les procédés de gestion des documents d'archives (compréhensibilité des documents, accessibilité, utilisabilité, protection, reproduction à long terme et migration, interopérabilité). Siegrist (2007) fournit un excellent résumé de ces deux normes ISO.

Qu'est-ce que les métadonnées? Partant de l'expression «des données sur les données», on peut définir les métadonnées de la manière suivante:

Les métadonnées sont des informations structurées, qui décrivent, expliquent ou retracent une ressource informative. A défaut, elles peuvent en faciliter la gestion (NISO¹⁵, Understanding Metadata); cette ressource peut être une publication, un rapport, un journal, un site web sur un thème ou un service particulier, ou un objet numérique tel qu'une image (National Library of Australia)¹⁶.

Dans les bibliothèques, les métadonnées sont généralement utilisées de la manière suivante:

Schéma formel de description d'une ressource, applicable à tous les types d'objets, aussi bien numériques qu'analogiques (NISO, 2004, 1)¹⁷.

Schéma formel de description d'une ressource, applicable à tous les types d'objets, aussi bien numériques qu'analogiques (NISO, 2004, 1).

Ces définitions établissent d'une part la fonction descriptive et d'autre part l'aspect technique des métadonnées, qui contiennent des informations sur d'autres données, ou données primaires. Ces descriptions contiennent des éléments d'information, qui ne peuvent généralement pas être obtenus directement dans les données primaires¹⁸. Le thème des métadonnées est toujours un sujet de recherche et doit par conséquent être considéré comme un ouvrage encore sur le métier, ou work in progress. Par ailleurs, les normes et les schémas en la matière sont développés constamment. Pour l'instant, il manque une norme de jure pour les images numériques¹⁹, comme c'est le cas pour l'organisation et la gestion des documents d'archives (Records Management)²⁰. Ces normes et modèles constituent des descriptions et des lignes directrices, qui doivent encore être explicitées lorsqu'elles sont appliquées. La mise en œuvre technique concrète, qu'elles n'abordent généralement pas, relève de l'institution qui en est responsable.

Concernant la caractérisation d'objets numériques, il existe plusieurs modèles de métadonnées, qui servent de standards de facto (p. ex. Dublin Core, EAD, MARCXML, etc.). L'échange d'informations n'est pas toujours simple. L'interopérabilité souvent citée implique un investissement important et ne peut être obtenue qu'au moyen de connaissances techniques solides, même si des «tableaux de concordance» entre certains standards sont définis dans le domaine de la gestion des documents d'archives²¹. Une vue d'ensemble des standards les plus répandus est disponible en annexe 10.3.

²¹ P. ex. entre ISAD(G), EAD, MAR-CXML, PREMIS etc. Les bibliothèques du MIT (MIT-Libraries) donnent un aperçu des mappages courants de métadonnées: <http://libraries.mit.edu/guides/subjects/metadata/mappings.html>. Harpring et al. (2006) donne une vue d'ensemble globale des standards les plus connus, étayée par des exemples.

²² Les différents types sont en général classés d'après leur fonctionnalité. Les frontières ne sont pas toujours évidentes. Par exemple, le numéro ISBN d'un livre peut être classé soit dans la catégorie descriptive, soit dans la catégorie administrative.

²³ Celles-ci sont souvent décrites comme «resource discovery metadata».

²⁴ C'est pourquoi on parle de «métadonnées insérées» (embedded metadata) par opposition aux métadonnées «stand-alone» (qui ne sont pas imbriquées, qui sont indépendantes).

3.1 CLASSIFICATION

Les métadonnées sont réparties en différentes catégories selon leur fonction. Un classement largement répandu (Hurley, et al., 1999) se limite à trois catégories: les métadonnées descriptives, structurelles et administratives²². Un autre modèle prévoit cinq catégories: description (resource description), recherche (information retrieval), administration (management), possession (ownership and authenticity), et interopérabilité (interoperability) (Haynes 2004, pp. 15 à 17). On critique généralement le fait que dans la plupart des classifications, trop peu d'importance est accordée au contexte d'utilisation (Day 2005, p. 21). En outre, dans ces deux modèles, les métadonnées techniques ne sont pas mentionnées explicitement. Elles font pourtant partie des objets numériques, même si leur définition se fait en règle générale automatiquement et en arrière-plan. Dans les paragraphes suivants, nous aborderons donc de plus près les métadonnées descriptives, structurelles, administratives et techniques, dans le contexte des fichiers images. Etant donné qu'il est souvent fait référence à ces quatre catégories, nous partons du principe qu'il existe un consensus en la matière et que cette classification est durable.

Les métadonnées descriptives (descriptive metadata) servent à l'identification et à la recherche d'objets numériques, mais aussi analogiques²³. Cette catégorie compte entre autres des éléments comme le titre, l'auteur, le thème, des mots-clés, ainsi que d'autres éléments descriptifs tels que les propriétés physiques d'un objet (taille, média et état). Souvent, les métadonnées descriptives sont administrées de manière externe, p. ex. dans des banques de données ou des fichiers textes, stockés avec l'objet. Pour les images numériques, ces informations doivent en plus être insérées²⁴ dans le fichier image, et ne pas être isolées dans un second fichier. Le lien entre les données et les métadonnées est conservé et on veille facilement à ce que les métadonnées soient mises à jour en même temps que l'objet (NISO, 2004). Pour les images numériques, il est donc recommandé qu'un identifiant de l'objet (ID objet: qui identifie l'objet sans équivoque), le titre, l'auteur et l'éditeur soient intégrés au fichier (Puglia et al., 2004).

Les métadonnées structurelles décrivent les relations entre les différents composants d'une ressource numérique, réunissent les objets individuels en une unité et montrent comment celle-ci est articulée (table des matières, chapitres, etc.). Les métadonnées structurelles décrivent également la structure d'objets complexes (objets composés), formés de plusieurs objets individuels souvent très différents. Ces informations sont mises à contribution p. ex. lors de l'utilisation de ressources numériques pour l'affichage et pour la navigation entre les différents objets.

²⁵ Une liste de mots-clés peut par exemple faire référence à l'emplacement physique d'un objet. Il n'est pas rare que des distances importantes séparent l'information relative à l'objet de l'objet lui-même. L'utilisation immédiate en est rendue plus difficile.

Les métadonnées administratives servent surtout à la gestion des données et englobent des éléments tels que le loueur, le collaborateur, l'éditeur, des liens (p. ex. vers la table des matières), le numéro ISBN (pour les livres), la date de remise (p. ex. pour les dissertations), la date de mise en ligne, la date de modification, «disponible jusqu'à», la référence, le site Web dont l'objet fait partie, la version, le copyright, les droits d'accès, le processus de création, le format du fichier, la source, etc. Les métadonnées administratives peuvent aussi contenir des informations concernant les droits, la reproduction, les critères de sélection, les registres, etc., ou des renseignements relatifs à la facturation et des conventions pour la conservation à long terme des objets numériques. Les frontières avec les autres catégories sont souvent floues.

En tant que forme particulière des métadonnées administratives, les métadonnées techniques assurent et documentent la qualité et le caractère intact d'un objet numérique. Ces informations permettent de vérifier l'intégrité des objets. Les métadonnées techniques comprennent des éléments tels que le format, la ressource, la taille, le matériel, le facteur de compression (pour les images numériques), etc. De plus, les métadonnées techniques peuvent contenir des informations spécifiques au matériel ou au logiciel, mais également gérer des informations propres au format, comme des indications relatives à la migration de format. Dans le domaine des images numériques, les données des appareils photo appartiennent à cette catégorie (données Exif). Ces dernières sont inscrites automatiquement dans le fichier image lors de la prise de la photo.

3.2 CONCLUSIONS

Tandis que dans le monde analogique, les métadonnées sont souvent séparées de l'objet physique décrit²⁵, elles peuvent, dans le contexte numérique, aussi bien être conservées à côté de l'objet (p. ex. dans une banque de données d'inventaire) qu'intégrées dans l'objet lui-même. L'exploitation de ces deux possibilités est une mesure de sécurité adaptée. La description de l'objet requiert le choix d'un standard de métadonnées connu et largement répandu, qui supporte les catégories de métadonnées nécessaires et garantit la reproductibilité de l'objet numérique, de son utilisation active à son archivage numérique à long terme.

4 STANDARDS DE MÉTADONNÉES

²⁶ cf. Woodley, M. S. (2005).

²⁷ TASI, *Metadata Standards and Interoperability* (cf. bibliographie). La TASI met à disposition d'autres «Advice Papers» à disposition gratuitement.

²⁸ Il s'agit du standard ISO 15836. D'autres standards sont en cours de développement.

²⁹ Une vue d'ensemble des standards mentionnés est présentée en annexe.

³⁰ Tous développés par la Library of Congress, voir la liste des abréviations.

³¹ Ces schémas, qui ne sont pas très répandus, ne sont pas de réels acronymes: cf. SPECTRUM (*The UK Museum Documentation Standard*) et MIDAS (*UK Historic Environment Data Standard*).

³² Dans le domaine de la gestion des documents d'archives, il n'est pas rare qu'un objet soit accompagné de trois schémas (METS, EAD, PREMIS).

³³ Un «nivellement par le bas», également appelé «dumbing-down», peut, par exemple, consister à réunir deux éléments comme le prénom et le nom en un seul élément «personne».

Outre la description des objets numériques en elle-même, les standards et les modèles de métadonnées soutiennent un échange d'information efficace et pertinent via différents ordinateurs, réseaux et systèmes d'exploitation²⁶. Cette propriété, appelée interopérabilité, vise à faciliter l'utilisation conjointe de plusieurs groupes de données et collections différents. Les notions de standard et de schéma sont généralement utilisées sans distinction. Elles sont la plupart du temps employées comme des synonymes et ne se rapportent en général qu'à l'organisation formelle des éléments de métadonnées.

Un standard est une norme ou une réglementation approuvée par un groupe normatif international et revêt un caractère obligatoire (standard de jure). Dans le présent contexte, un schéma est une norme née de sa propagation (standard de facto).

La description des objets numériques est régie par de nombreux standards (ou schémas), qui ne sont pas seulement difficiles à sélectionner parce que leur désignation se résume à un acronyme. Il est également ardu d'en donner une vue d'ensemble compréhensible, qui peut au premier abord effrayer les profanes, en dépit de bonnes explications. A cet égard, la classification du TASI est une introduction pertinente, qui donne un bon aperçu²⁷.

Comme la plupart des aperçus, celui du TASI mentionne Dublin Core comme le premier standard de référence et schéma de description. Dublin Core, qui a, à l'origine, été conçu pour la description de sites Web, est le standard le plus utilisé, ne serait-ce qu'en raison de sa simplicité; c'est également l'un des rares standards certifiés ISO²⁸. Parmi les autres «schémas d'images» cités ensuite, on trouve VRA Core, CDWA et SEPIADES²⁹.

Parmi les standards «spécifiques aux communautés», on citera MARC, MARCXML et MODS³⁰ pour les bibliothèques, SPECTRUM, MIDAS³¹ pour les musées (presque exclusivement dans un contexte anglophone), ISAD(G), EAD, DACS pour les archives et TEI pour la désignation de textes scientifiques.

Les schémas spécifiques aux tâches comprennent NISO (métadonnées techniques pour les images numériques), PREMIS (archivage à long terme de données numériques), METS (conteneur et paquet pour d'autres schémas), LOM (ressources didactiques) et OAI-PMH (interopérabilité avec d'autres standards).

Exif, IPTC et XMP font partie des «métadonnées d'images». Ces informations sont inscrites directement dans le fichier image (TIFF et JPEG) et peuvent ainsi être transmises avec l'image sans que cela ne nécessite d'investissement supplémentaire.

La quintessence de cette classification et énumération prouve qu'un standard unique ne serait évidemment pas suffisant pour décrire un objet numérique dans son intégralité. Espérons qu'à l'avenir, des simplifications et le «nivellement par le bas» de certains éléments puissent alléger les descriptions³³.

³⁴ «PREMIS defines “preservation metadata” as information a repository uses to support the digital preservation process [...]»

³⁵ «[...] any metadata absolutely required under any circumstances [...] Core does not necessarily mean mandatory, and some semantic units were designated as optional when exceptional cases were apparent.»

La plupart des standards de métadonnées connus continuent à se limiter à la description d’objets numériques, destinée avant tout à faciliter la recherche ciblée. Cette manière purement synchrone de considérer les choses, qui est tout à fait justifiée dans le travail quotidien avec des objets numériques, risque, dans un avenir proche, de conduire non seulement à des difficultés techniques, mais aussi à des problèmes d’interprétation du contexte.

Depuis quelques années, les développements dans le domaine des métadonnées destinées à la conservation à long terme (preservation metadata) ont été suivis avec un intérêt croissant. Outre Cedars et PADI, c’est surtout le groupe de travail PREMIS qui a traité ce thème de manière plus détaillée. Dans son rapport final, qui présente également un dictionnaire des données (Data Dictionary) détaillé, PREMIS définit les métadonnées destinées à la conservation à long terme de la façon suivante:

Informations utilisées par un répertoire [des archives numériques] pour soutenir le processus d’archivage numérique (PREMIS, 2005)³⁴.

Cette vision fonctionnelle englobe les différents types de métadonnées (descriptives, structurelles, etc.).

La notion de métadonnées de base est définie de manière un peu plus diffuse:

Métadonnées requises ou devant être disponibles en toutes circonstances; «de base» ne signifie pas forcément «obligatoire».

Cette contradiction est expliquée par le fait que dans certains cas de toute évidence exceptionnels, des «unités sémantiques» (donc des champs de métadonnées) peuvent être facultatives³⁵. Cette affirmation sortie de son contexte prouve une fois de plus que de nombreuses zones d’ombre subsistent et qu’il faut continuer à chercher des solutions praticables.

Par ailleurs, de nombreuses décisions sont pour l’instant reportées en raison de la complexité du sujet, dans l’espoir que de futures solutions commerciales résolvent le problème dans son ensemble et permettent de trouver un dénominateur commun aux aspects relevant de la technique et du contenu.

³⁶ C'est pourquoi il est recommandé d'utiliser des logiciels commerciaux (p. ex. Photoshop) pour les conversions.

4.1 CONCLUSIONS RELATIVES AUX MÉTADONNÉES D'IMAGES

Il existe trois types de métadonnées qui peuvent être utilisées et échangées dans des fichiers images (TIFF, JPEG, etc.) (Important pour les images: il ne faut pas perdre les métadonnées lors des conversions de format!)³⁶.

1. Exif: Informations techniques sur les images prises par un appareil photo ou un scanner. Ces données sont générées automatiquement et passent quasiment inaperçues pour l'utilisateur, qui ne les modifie généralement pas. Le champ principal est celui de la date de création de l'image (Creation Date).
2. IPTC: Possibilité d'intégrer des descriptions simples aux fichiers images. Avantage: métadonnées à large diffusion, supportées par de nombreux programmes de traitement de l'image. Il existe de nombreux paquets de logiciels de qualité capables de traiter de grandes quantités d'images. Inconvénient: les champs sont conçus pour la photographie de presse et sont souvent exploités, pour sauvegarder d'autres informations que celles prévues, ce qui peut être déconcertant (cf. exemple d'application). Un autre inconvénient est que les champs sont relativement courts, ce qui ne permet pas, par exemple, de décrire l'image en détail.
3. XMP: Pour pallier les inconvénients d'IPTC, Adobe a défini XMP avec de nombreux ajouts. XMP ne fonctionne bien qu'avec les logiciels Adobe, même si de plus en plus de paquets de logiciels supportent cette «plate-forme». Il reste toutefois encore beaucoup d'opérations à effectuer manuellement et il n'existe pour l'instant aucun lien vers une banque de données, ce qui devrait néanmoins être possible à l'avenir.

5 EXEMPLE D'APPLICATION

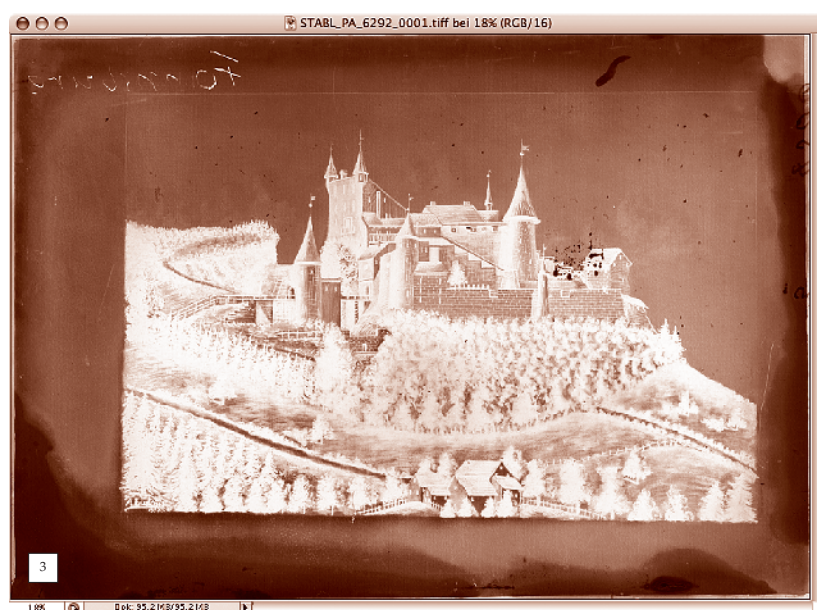
³⁷ Le Imaging & Media Lab remercie les archives d'Etat de Bâle-Campagne de lui avoir autorisé à utiliser une partie de ses images à des fins d'illustration.

³⁸ Cette information se trouvait dans le fichier Excel fourni, sous la rubrique «Source».

L'exemple ci-dessous montre comment compléter de manière simple des images numériques par des métadonnées. Les images utilisées à cet effet ont été envoyées à l'IML (Imaging & Media Lab) par les archives d'Etat de Bâle-Campagne en vue de leur numérisation. Les archives ont livré les originaux analogiques accompagnés des descriptions des images (dans un fichier Excel)³⁷.

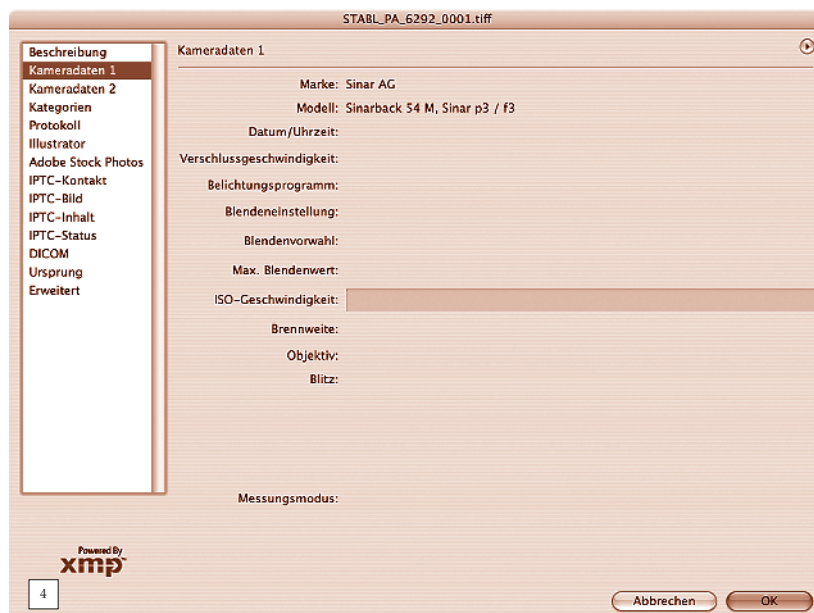
Etape 1

Les images sont numérisées au moyen d'un appareil photo numérique SINAR et sont immédiatement renommées avec le nom de fichier correct, correspondant à la convention de dénomination prévue en la matière³⁸.



Les métadonnées techniques (données Exif) ont été automatiquement saisies dans le fichier lors de la numérisation. Ces informations peuvent par exemple être affichées dans Photoshop. Dans cet exemple, seuls la marque et le modèle d'appareil photo sont indiqués; normalement, davantage d'informations s'affichent pour les prises de vue à partir d'un appareil photo numérique.

³⁹ Le programme CaptionWriter de la société Comnet Software permet d'importer des descriptions IPTC pour une collection complète d'images. Adobe Bridge ne permet pour l'instant qu'une saisie semi-automatique des descriptions dans les champs IPTC après sélection des images souhaitées, ou l'importation de métadonnées pour une seule image.



Etape 2

Une fois la numérisation terminée, des versions JPEG haute résolution sont créées à partir du document TIFF. Celles-ci peuvent être utilisées comme copies de travail ou pour la diffusion. Pendant ce temps, le fichier maître TIFF («l'original») reste inchangé et peut être utilisé pour créer d'autres copies.

Etape 3

Après la production des fichiers JPEG, on passe à la définition des métadonnées descriptives, que le donneur d'ouvrage a livrées dans un fichier Excel. Il est possible qu'elles proviennent d'une banque de données. Les noms des champs correspondent déjà aux noms des champs IPTC. Grâce à l'attribution nette du nom de fichier (p. ex. STABL_PA_6292_0001.TIFF dans le champ ObjectName, voir image à la page suivante), il est possible de procéder à une saisie automatique des métadonnées. Pour un travail portant sur plus de 1000 images, cela permet de gagner beaucoup de temps. En outre, le nombre d'erreurs qui résultent inévitablement d'une saisie manuelle n'est pas seulement réduit, il passe carrément à zéro. Pour la saisie automatique, on utilise le programme CaptionWriter³⁹.

⁴⁰ Cette limitation devrait être supprimée à l'avenir avec IPTC4XMP, mais cela suppose que le logiciel utilisé supporte le cadre de métadonnées XMP d'Adobe (cf. <http://www.iptc.org/IPTC4XMP/>).

ObjectName	CreateDate	City	Province	Country	SpecialInstructions	Byline	Headline
STABL_PA_6292_0001	1895	Ormalingen Far	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.001 / 1(a)	Farnsburg, ab B	
STABL_PA_6292_0002	undatiert	Ormalingen Far	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.002 / 1(b)	Wachsfisuren-R	
STABL_PA_6292_0003	undatiert	Ormalingen Far	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.003 / 1(c)	Wachsfisuren-R	
STABL_PA_6292_0004	1891	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.004 / 2	Liestal, Rathau	
STABL_PA_6292_0005	01.06.1896	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.005 / 3	Liestal, Kasern	
STABL_PA_6292_0006	18.06.1934	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.006 / 4	Liestal, Kasern	
STABL_PA_6292_0007	1896	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.007 / 5	Liestal, Kasern	
STABL_PA_6292_0008	1864	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.008 / 6	Liestal, Kasern	
STABL_PA_6292_0009	30.04.1944	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.009 / 7	Liestal, Kasern	
STABL_PA_6292_0010	02.10.1929	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.010 / 8	Liestal, Burgst	
STABL_PA_6292_0011	1910	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.011 / 9	Liestal, Burgst	
STABL_PA_6292_0012	1910	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.012 / 10(a)	Liestal, Burgst	
STABL_PA_6292_0013	1910	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.013 / 10(b)	Liestal, Burgst	
STABL_PA_6292_0014	1911	Liestal	BL	Schweiz	Reproduktion ab Glasnegativ KM 00.014 / 11	Liestal, Burgst	

La définition a été faite séparément pour la version TIFF et la version JPEG. Cette procédure a une raison: il n'est pas rare que lors des conversions de format, des informations contenues dans les champs IPTC soient perdues (p. ex. de TIFF en JPEG). En outre, avec IPTC, le nombre des champs de description est fixé, la longueur des champs est limitée et peut être trop courte selon l'application⁴⁰. Ces informations peuvent également être visualisées dans Photoshop ou dans de nombreux autres programmes de traitement de l'image.

STABL_PA_6292_0001.tiff

Beschreibung

- Kameradaten 1
- Kameradaten 2
- Kategorien
- Protokoll
- Illustrator
- Adobe Stock Photos
- IPTC-Kontakt
- IPTC-Bild
- IPTC-Inhalt**
- IPTC-Status
- DICOM
- Ursprung
- Erweitert

IPTC-Inhalt

Verwenden Sie dieses Fenster, um den visuellen Inhalt des Bildes zu beschreiben.

Überschrift Farnsburg, ab Bild A. Lüdin

Beschreibung Schloss Farnsburg, Reproduktion eines Bildes von A. Lüdin, 1895

Stichwörter*

IPTC-Subject Code*

Themencodes sind unter <http://www.newscodes.org> definiert.

Verfasser der Beschreibung

* Mehrere Werte können durch ein Komma oder Semikolon getrennt werden.

Powered By **xmp**

6

Abbrechen OK

⁵ Fichier Excel contenant les métadonnées IPTC

⁶ Contenu IPTC: STABL_PA_6292_0001.TIFF

STABL_PA_6292_0001.tiff

Beschreibung
 Kameradaten 1
 Kameradaten 2
 Kategorien
 Protokoll
 Illustrator
 Adobe Stock Photos
 IPTC-Kontakt
IPTC-Bild
 IPTC-Inhalt
 IPTC-Status
 DICOM
 Ursprung
 Erweitert

IPTC-Bild

Verwenden Sie dieses Fenster zur Eingabe formaler beschreibender Informationen über das Bild.

Erstellungsdatum

Genre

IPTC-Scene Code*

Szenenwerte sind unter <http://www.newscodes.org> definiert.

Ort des Motivs

Stadt/Ort

Bundesland/Kanton

Land

ISO-Landescode

Ländercodes bestehen aus 2 oder 3 Buchstaben gemäß Definition in der Norm ISO 3166.

* Mehrere Werte können durch ein Komma oder Semikolon getrennt werden.

7 

Abbrechen OK

STABL_PA_6292_0001.tiff

Beschreibung
 Kameradaten 1
 Kameradaten 2
 Kategorien
 Protokoll
 Illustrator
 Adobe Stock Photos
 IPTC-Kontakt
 IPTC-Bild
 IPTC-Inhalt
IPTC-Status
 DICOM
 Ursprung
 Erweitert

IPTC-Status

Verwenden Sie dieses Fenster, um Workflow- und Copyright-Informationen aufzuzeichnen.

Titel

Jobkennung

Anweisungen

Anbieter

Quelle

Copyright-Vermerk

Nutzungsbedingungen

8 

Abbrechen OK

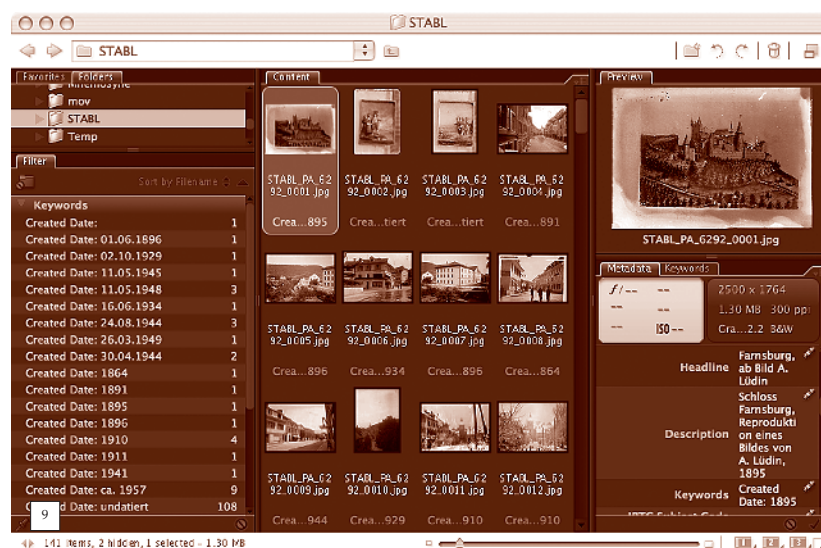
⁴¹ Des exemples de logiciels de gestion des images sont disponibles sur la page Web du TASI «Systems for Managing Image Collections»: <http://www.tasi.ac.uk/advise/delivering/ims2.html>.

⁴² L'exemple doit aussi montrer qu'il est difficile de définir des mappages (mises en concordance) entre les schémas qui ont été conçus pour des applications différentes, et qu'il n'existe souvent pas de concordances.

⁴³ Exemples pratiques de contenus de champs qui ne correspondent pas aux descriptions des champs d'origine.

⁴⁴ Il faudrait déterminer si cette information devrait être définie comme <subject> dans le titre ou dans la partie descriptive.

On peut obtenir un aperçu rapide de l'objet de l'image par exemple avec Adobe Bridge.⁴¹



Etape 4

Après la création des copies de travail JPEG et la saisie des métadonnées, les images sont transmises au donneur d'ouvrage, accompagnées des originaux analogiques, et les images numériques sont préparées pour l'archivage à long terme, qui sera traité au chapitre 7.

Cet exemple d'application montre uniquement comment les métadonnées sont inscrites dans les fichiers images. Il est également possible de gérer des méta-informations dans un fichier externe. Le tableau ci-dessous montre dans les grandes lignes comment les mêmes informations pourraient être saisies par exemple dans Dublin Core et EAD⁴²:

IPTC	Dublin Core	EAD	Exemple d'information
Titre (Title)	dc.title	<unititle>	STABL_PA_6292_0001 ⁴³
Description (Headline)	dc.subject	...<subject> ⁴⁴	Schloss Farnsburg ...
Auteur (Credit)	dc.creator	<author>	Seiler-Schaub
Mots-clés (Keywords)	dc.description?	?	Created Date: 1895?
...	...	...	...

6 CHAMPS PRINCIPAUX

⁴⁵ cf. <http://www.iptc.org/IPTC4XMP/>.

⁴⁶ Avec cette sélection limitée d'attributs, le problème de l'interopérabilité refait surface. La sémantique des différents champs peut diverger, ce qui risque de poser des problèmes d'interprétation en cas de transfert ultérieur dans un nouveau schéma.

⁴⁷ Dans le cas présent, on ne sait pas de quelle date on parle (actuelle ou historique?).

Dans le domaine des images numériques, la définition des métadonnées IPTC reste une procédure à recommander. Même si les métadonnées peuvent être gérées dans une application externe (p. ex. dans une banque de données), il est préférable de placer les mêmes informations également dans les fichiers images. Si à l'avenir la référence entre l'application et l'image devait disparaître, le contenu et le contexte de cette dernière pourraient être reconstitués au moyen des descriptions intégrées à l'image. Depuis 2004, IPTC et Adobe travaillent sur un schéma commun IPTC pour XMP (IPTC4XMP), qui synchronise dans les deux sens les informations entre les champs standard d'IPTC (version 1.0) et le nouveau schéma basé sur XML⁴⁵. La définition des métadonnées IPTC reste ainsi une manière de procéder judicieuse et valable à l'avenir.

Les recommandations des directives sur les métadonnées descriptives de RLG (RLG Descriptive Metadata Guidelines, 2005, cf. annexe) peuvent servir de base pour la définition des champs principaux. Dans le domaine des images numériques, les champs suivants permettent de décrire le contexte de manière étendue:

	RLG (proposition)	IPTC	Dublin Core
Identification	Unique ID	Object Name?	dc.identifier ⁴⁶
Description	Work Description	Caption	dc.description
Collection	Collection Title	Headline	dc.title
Thème	Collection Subjects	Category	dc.subject
Mots clés	?	Keywords	dc.subject
Créateur / auteur	«Creator»	Credit?	dc.creator
Date	Date Range?	Create Date	dc.date? dc.coverage? ⁴⁷

En principe, les champs principaux devraient se limiter à un minimum significatif. Les descriptions «approfondies», contenant de nombreux attributs, comportent le risque que, par manque de temps, on se contente de ne remplir que les champs «les plus importants».

7 ARCHIVAGE NUMÉRIQUE

Au terme des opérations numériques (numérisation, définition des métadonnées, création de copies de travail) se pose la question de la conservation à long terme des données numériques fraîchement créées. Le principal problème réside dans la sauvegarde d'informations numériques, qui est décrite à l'aide du modèle par couches simplifié suivant (Gschwind, 2007):

Informations, métadonnées	Dossiers, documents, photos, courriels, etc.
Logiciel, format des fichiers	Programmes d'application, fichiers
Matériel, support des données	Processeur, mémoire, bande magnétique, disques durs optiques et magnétiques (hard disks)

La couche la plus basse (couche matérielle) est extrêmement instable et a une durée de vie réduite, pour diverses raisons, les problèmes principaux étant l'évolution technologique, les supports de données obsolètes, l'usure des supports de données, etc.

Cette instabilité se répercute sur la couche intermédiaire (couche logicielle), p. ex. en cas de nouveaux formats. Le changement se fait plus lentement, mais pose tout autant de problèmes.

La couche supérieure (couche de l'information) ne relève en principe pas du pur problème numérique. Cependant, si la couche inférieure fait défaut, les fichiers et les métadonnées perdent toute pertinence et l'information est définitivement perdue.

Le problème de base de l'archivage durable des données numériques réside dans l'enregistrement des informations. À l'aide des ordinateurs, les informations sont converties en codes binaires et enregistrées sur des supports de données numériques (conservation du flux binaire). S'il est manipulé correctement, le code binaire peut être copié à volonté sans perte d'information. Pour les raisons mentionnées plus haut, il doit être migré à intervalles réguliers, en général tous les cinq ans au plus. Le résultat de ce processus est un échange complet de la technologie, au cours duquel les informations sont copiées sur de nouveaux supports de données. Cela doit se passer sans exception et sans interruption. Une migration engendre des coûts fixes élevés et nécessite des connaissances techniques suffisantes pour qu'aucune donnée ne soit perdue.

Le processus d'archivage expliqué ci-dessous est le résultat de plusieurs années de recherches et d'essais pratiques du laboratoire Imaging & Media Lab dans le domaine de l'archivage numérique des images. La procédure peut être utilisée aussi bien pour un premier archivage que pour les archivages ultérieurs et les migrations. Il s'agit d'une manière de procéder générique, qui peut être adaptée aux besoins spécifiques.

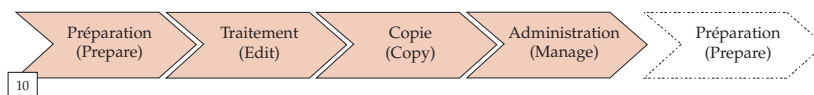
⁴⁸ Une description détaillée du processus d'archivage est disponible sur le site Web de l'IML, à l'adresse: <http://www.iml.unibas.ch/index.php?content=28&news=1>.

⁴⁹ Correspond en principe à un paquet de livraison dans le schéma OAIS: Submission Information Package (SIP), cf. OAIS Reference Model dans la bibliographie.

⁵⁰ cf. Message Digest Algorithm 5 (MD5): <http://fr.wikipedia.org/wiki/MD5>.

7.1 PROCESSUS D'ARCHIVAGE

Le processus d'archivage se compose de quatre parties: la préparation, le traitement, la copie et l'administration⁴⁸:



7.1.1 PRÉPARATION

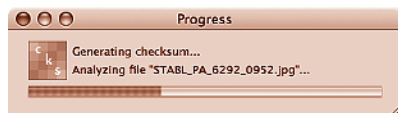
La préparation commence avec l'enregistrement des données numériques qui viennent d'être créées ou qui sont produites à partir d'une source analogique. Cette partie du processus a déjà été illustrée dans l'exemple d'application (chapitre 5, étape 1). Il est recommandé de définir une structure d'archives compréhensible par les utilisateurs des données, pour toutes les collections de données (hiérarchie de dossiers ou liste). Les étapes suivantes font, entre autres, partie de la préparation:

- Evaluation et sélection
- Définition d'un paquet de données d'archivage (liste, nombre de fichiers)⁴⁹
- Détermination de la quantité de données (influe sur les coûts, cf. sélection)
- Documentation de la structure de la liste
- Choix du support d'archivage
- Vérification de l'infrastructure d'archivage
- Fixation du moment d'archivage (utilisation limitée de l'infrastructure de TI)

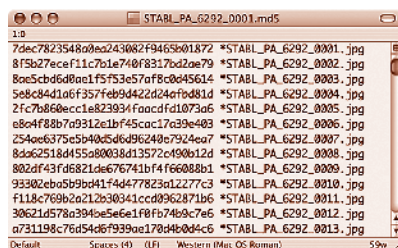
7.1.2 TRAITEMENT

Cette partie consiste, d'une part, à définir les métadonnées nécessaires à l'archivage (métadonnées descriptives et administratives; cf. exemple d'application, ch. 5, étape 3) et, d'autre part, à calculer des sommes de contrôle, ou empreintes, pour tous les fichiers et à les mettre dans un fichier texte, qui est archivé avec les fichiers images. Les principales étapes sont les suivantes:

- récapituler les fichiers (thème, contenu de la liste, nombre);
- définir et traiter les métadonnées IPTC (individuellement, par groupes, automatiquement);
- vérifier les descriptions (contrôle de la qualité, intégralité, schéma de classification?);
- choisir un algorithme de calcul (p. ex. MD5⁵⁰);
- calculer les sommes de contrôle pour les fichiers sélectionnés ou le paquet d'archives;
- placer le fichier de la somme de contrôle dans la même liste (base de comparaison pour les vérifications ultérieures de l'intégrité, archivée avec les données images).



11



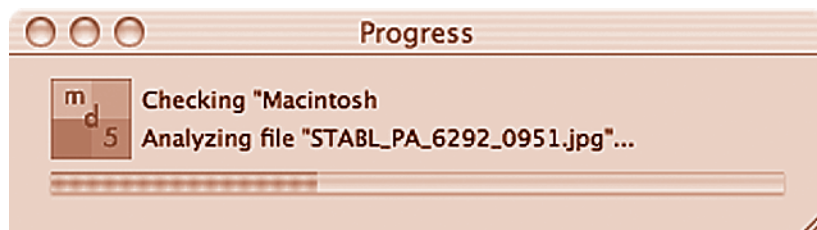
12

Les sommes de contrôle permettent une vérification simple de l'intégrité des données. Grâce aux sommes de contrôle, une comparaison permet de constater si une transmission ou un enregistrement de données s'est déroulé correctement (p. ex. sans erreur de transmission).

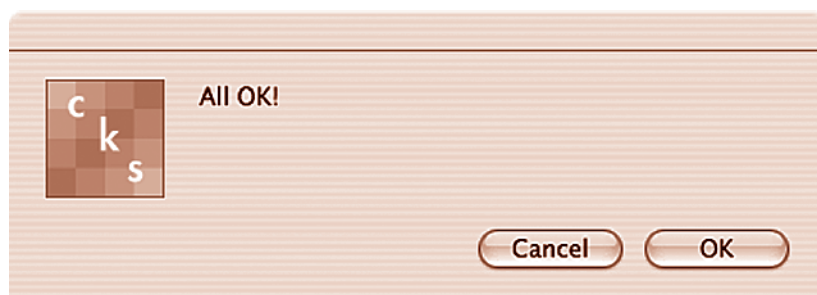
1.3 COPIE

Cette étape consiste à transférer les fichiers à archiver dans les archives numériques, puis à recalculer ensuite les sommes de contrôle et à les comparer avec celles qui se trouvent dans le fichier texte archivé (cf. deuxième étape, Traitement). Les opérations principales de cette phase sont les suivantes:

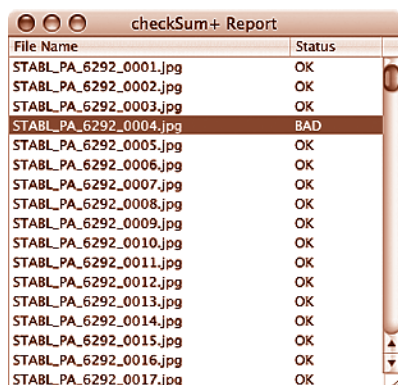
- copier les fichiers dans les archives numériques;
- créer au moins trois copies (qui seront plus tard conservées à des endroits différents);
- après chaque copie, recalculer les sommes de contrôle dans les archives et les comparer;



13



14



15

- en cas d'erreurs, en déterminer la cause (erreur de transmission, média, etc.);
- corriger les erreurs immédiatement!

Si les sommes de contrôle des fichiers originaux sont identiques à celles des archives, la copie archivée correspond au fichier d'origine (contrôle d'intégrité réussi). Si les sommes ne sont pas identiques, c'est que l'on est en présence d'une erreur. Cette dernière peut avoir plusieurs causes: transmission incomplète, média d'archivage endommagé, erreur de manipulation, etc. Dans ce cas, le fichier contenant l'erreur doit être remplacé. A cet égard, on ne soulignera jamais assez l'importance de sauvegarder les données à double.

11 Calcul de la somme de contrôle (p. ex. avec CheckSum+)

12 Contenu du fichier de la somme de contrôle

13 Comparaison de sommes de contrôle

⁵¹ Le processus d'archivage présenté ici ne couvre que les deux premiers composants (livraison et archivage). L'utilisation a été laissée de côté intentionnellement, car elle ne fait en principe pas partie de l'archivage.

⁵² Des solutions logicielles sont développées pour la gestion d'archives numériques. Dans le domaine de l'Open-Source, Fedora (<http://fedoraproject.org/>) et Dspace (<http://www.dspace.org/>) suscitent de plus en plus l'intérêt.

7.1.4 ADMINISTRATION

Après le transfert des fichiers dans les archives numériques, la partie principalement technique de l'archivage est terminée. L'administration des archives englobe différentes tâches de conduite, qui doivent être reproductibles en tout temps et ne devraient pas dépendre d'une personne déterminée. L'administration d'archives numériques comprend, entre autres, les activités suivantes:

- documentation des données archivées (quelles données, emplacement où elles sont stockées, etc.);
- classement des données (types de données, accès);
- planification et exécution de procédures de relecture périodique (p. ex. tous les six mois, vérifier que le média est lisible, recalculer et comparer les sommes de contrôle);
- planifier la première ou la prochaine migration;
- budgétiser à temps le projet de migration;
- définir le lieu de conservation (à l'intérieur ou à l'extérieur, selon les critères prix/sécurité/accessibilité);
- planifier et budgétiser la formation et le perfectionnement du personnel.

En fonction des volumes de données, la sauvegarde peut prendre des heures, voire des jours. C'est pourquoi la planification de la procédure ad hoc est très importante. Il faut en outre penser qu'un processus d'archivage correct implique l'enregistrement de données sur des supports redondants. Le temps nécessaire pour sauvegarder les données doit par conséquent être multiplié par le nombre de copies à créer. Une fois les données sauvegardées, les supports de données et la documentation qui les accompagne (cf. Préparation) doivent être stockés en divers endroits, de préférence dans des bâtiments différents, suffisamment éloignés géographiquement les uns des autres. Depuis quelque temps, il est recommandé de gérer les archives numériques selon le modèle de référence OAIS. Ce modèle prévoit trois composants principaux: la livraison sous forme de SIP (Submission Information Package), l'archivage sous forme d'AIP (Archive Information Package) et l'utilisation, sous forme de DIP (Dissemination Information Package)⁵¹. Le modèle de référence OAIS ne fait qu'établir les conditions-cadres. La mise en œuvre effective des fonctions de livraison, d'archivage et d'utilisation est laissée aux bons soins des institutions d'archivage⁵². Des informations plus approfondies sont disponibles dans les ouvrages de référence (cf. bibliographie).

7.2 CONCLUSIONS

Tandis qu'un premier archivage constitue un processus qui crée de la valeur, les migrations subséquentes sont surtout des processus de maintien de la valeur, engendrant des coûts récurrents. Il faut qu'au moment de la migration, les moyens financiers nécessaires à ce maintien, mais aussi à l'infrastructure requise, soient à disposition. S'il fallait renoncer à une migration en raison du manque de ressources financières, des données numériques pourraient rapidement disparaître. Il faut en être conscient et faire en sorte que des engagements financiers soient pris en conséquence.

¹⁴ Comparaison de sommes de contrôle en ordre (cf. p. 26)

¹⁵ Erreur dans les sommes de contrôle (cf. p. 26)

8 CONCLUSIONS FINALES

ET PERSPECTIVES

⁵³ Depuis 2003, le DCMI est une norme ISO (15836-2003) et depuis mai 2007, une norme NISO (Z39.85-2007).

⁵⁴ XMP d'Adobe insère des métainformations conformes aux éléments de Dublin Core.

Tandis que les procédures se consolident peu à peu dans le domaine de la numérisation, le problème des métadonnées reste complexe et est souvent réglé de manière insatisfaisante. Bien que de nombreux schémas existent (cf. annexe), la solution pour décrire et gérer les biens culturels numériques de façon homogène reste encore à trouver. Pourtant, la définition de métadonnées est indispensable pour que les objets numériques restent accessibles et saisissables à l'avenir. Cela s'applique en particulier aux images numériques, dont on ne peut pas déduire le contenu au moyen d'informations textuelles.

La définition des métadonnées dans l'image numérique elle-même présente quelques avantages. L'objet contient toutes les informations qu'on y enregistre et ne dépend en principe pas de l'appui d'un logiciel de gestion; cela ne signifie pas que l'administration externe des informations et du contexte ne procure que des inconvénients. Pour les tâches quotidiennes, un logiciel de gestion offre au contraire de nombreux avantages, comme une meilleure vue d'ensemble ou un accès plus rapide à un objet numérique. Par contre, l'insertion de métadonnées dans le fichier image simplifie le processus d'archivage, puisque lorsqu'on crée une copie, toutes les informations sont automatiquement transférées avec l'objet. Ce procédé a toutefois également des inconvénients: les champs Exif et IPTC peuvent être facilement modifiés ou effacés. C'est pourquoi il convient de prendre des mesures supplémentaires, pour protéger les images numériques, mais aussi les objets numériques en général, contre des manipulations involontaires.

En ce qui concerne les normes de métadonnées, Dublin Core fait toujours office de schéma de référence. Dublin Core existe depuis 1995 et se compose d'une liste bien structurée de 15 éléments, qui peuvent être intégrés simplement aux documents, tant aux images numériques qu'aux documents HTML par exemple. Le groupe Dublin Core Metadata Initiative s'engage en faveur de l'interopérabilité des métadonnées et constitue désormais aussi une norme de jure⁵³, à laquelle d'autres normes se réfèrent⁵⁴. Lorsque l'on choisit un schéma, il faut donc veiller à ce que l'interopérabilité avec Dublin Core soit garantie. Dublin Core peut être considéré comme une «lingua franca» parmi les normes relatives aux métadonnées. D'autres normes doivent pouvoir être simplifiées avec des versions Lite (p. ex. CDWA Lite, VRA Core, cf. annexe). On peut s'attendre à de nouvelles simplifications à l'avenir. A cet égard, limiter les champs principaux à une quantité que l'on puisse embrasser d'un regard est une méthode à retenir.

Dans le domaine de l'archivage numérique, la phase de sensibilisation appartient au passé. L'archivage numérique signifie l'enregistrement d'une situation déterminée sur un support de données, qui est ensuite conservé en lieu sûr sans être modifié, jusqu'à la prochaine migration. La procédure est clairement définie. Elle engendre également la nécessité d'un financement à long terme, qui permet de garantir la disponibilité des archives numériques. L'archivage numérique est un processus actif, qui ne tolère aucune interruption et requiert des activités systématiques.

La mise en œuvre de la procédure fixée dans ces directives constitue une première étape. Pour que les images numériques restent disponibles et saisissables également à l'avenir, il convient de formuler quelques recommandations du point de vue actuel; elles sont énumérées dans la liste de contrôle annexée (chap. 10.1). Les présentes directives reflètent la situation actuelle. Les technologies, les logiciels et les normes sont soumis à des changements permanents. Par conséquent, il sera nécessaire d'adapter les manières de procéder décrites dans cet ouvrage aux nouvelles conditions-cadres et circonstances.

9 BIBLIOGRAPHIE

*Etat:
Février 2008*

Day, M. (2005). DCC Digital Curation Manual, Instalment on Metadata:
<http://www.dcc.ac.uk/resource/curation-manual/chapters/metadata>

Deutsche Nationalbibliothek, Schlagwortnormdatei (fichier de normalisation des données):
<http://www.ddb.de/standardisierung/normdateien/swd.htm>

Gschwind, R. (2007). Braucht es in der Schweiz ein Kompetenzzentrum für die Langzeitarchivierung digitaler Daten? Berne: SAGW.

Harpring, P., Woodley, M. S., Gilliland, A. J. and Murtha, B. (2006). Metadata Standards Crosswalks in Pathways to Digital Information: Online Edition 2.1:
http://www.getty.edu/research/conducting_research/standards/intrometadata/crosswalks.html

Haynes, D. (2004). Metadata for Information Management and Retrieval. London: Facet.

Hurley, B. J., Price-Wilkin, J., Proffitt, M. and Besser, H. (1999). The Making of America II Testbed Project: a Digital Library Service Model, Washington, D.C.: Council on Library and Information Resources.

Margulies, S. (2007). Metadata and Digital Photography: in Peres, Michael R. (ed.). Focal Encyclopaedia of Photography, 4th. Ed., Elsevier Focal Press: 411-417.

Menne-Haritz, A. (2005). METS: Überblick und Anleitung (traduction en allemand de METS: An Overview & Tutorial):
<http://www.loc.gov/standards/mets/METSOverview.v2.html>

NISO, National Information Standards Organisation (2004). Understanding Metadata:
<http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>

OAIS, Reference Model for an Open Archival Information System. Consultative Committee for Space Data Systems:
<http://public.ccsds.org/publications/archive/650x0b1.pdf>

OCLC/RLG, Online Computer Library Center / Research Libraries Group (2005), Data Dictionary for Preservation Metadata, Final Report of the PREMIS Working Group (May 2005):
<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>

Puglia, S., Reed, J. and Rhodes, E. (2004). Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images:
<http://www.archives.gov/research/arc/digitizing-archival-materials.pdf>

RLG, Research Libraries Group (2005). Descriptive Metadata Guidelines for RLG Cultural Materials, RLG:
http://www.rlg.org/en/pdfs/RLG_desc_metadata.pdf

Rosenthaler, L. (2007). Digital Archiving: in Peres, Michael R. (ed.). Focal Encyclopaedia of Photography, 4th. Ed., Elsevier Focal Press: 359-364.

Siegrist, B. (2007). Organisatorische Voraussetzungen des betrieblichen Informationsmanagements unter Nutzung des IKT. Diplomarbeit an der Hochschule für Wirtschaft, Institut für Wirtschaftsinformatik: Olten.

TASI, Technical Advisory Service for Images, Metadata Standards and Interoperability, Advice Paper:
<http://www.tasi.ac.uk/advice/delivering/metadata-standards.html>

Woodley, M. S. (2005). Dublin Core Metadata Initiative Glossary, DCMI:
<http://dublincore.org/documents/usageguide/glossary.shtml>

Association des Archivistes suisses (AAS), Catalogue des normes archivistiques importantes utilisées en Suisse:
http://www.vsa-aas.org/uploads/media/CatalogueNormes_Version1-2_20090201fr.pdf

10 ANNEXE

La liste de contrôle ci-dessous résume les points principaux et est conçue à des fins d'orientation.

10.1 LISTE DE CONTRÔLE

Généralités	
Métadonnées	
Saisir des informations dans les champs IPTC pertinents des images numériques	
Indiquer la signification spéciale d'attributs (divergence par rapport à la dénomination d'origine) (en particulier pour les champs IPTC)	
Limiter les champs principaux à un minimum pouvant être embrassé d'un seul coup d'oeil	
...	
Schéma de métadonnées	
Utiliser Dublin Core comme schéma de référence ou le prendre en considération comme schéma d'orientation (interopérabilité, «concordance des champs» pour les changements de schéma)	
Suivre le développement des normes et schémas de métadonnées	
...	
Suivre l'évolution technologique (technology watch)	
Suivre l'évolution des médias de sauvegarde	
Suivre l'évolution des formats de fichiers	
Vérifier les migrations de formats (qualité, perte d'information, réversibilité, etc.)	
...	
Archivage numérique (processus)	
Préparation	
Evaluation et sélection	
Définir des paquets d'archives	
Déterminer les quantités de données	
Documenter la structure du répertoire	
Choisir le support d'archivage	
Vérifier l'infrastructure d'archivage	
Fixer le moment de l'archivage	
...	

Traitement	
Regrouper les fichiers (par thème, contenu du répertoire, nombre)	
Définir et saisir les métadonnées IPTC	
Vérifier les descriptions IPTC	
Calculer les sommes de contrôle pour les paquets d'archives	
Stocker le fichier des sommes de contrôle avec les données archivées (même répertoire)	
...	
Copie	
Copier les fichiers sur le média d'archivage (archive numérique)	
Créer au moins trois copies (redondance)	
Après chacune des opérations de copie, recalculer les sommes de contrôle et les comparer	
...	
Administration	
Garantir le financement à long terme	
Etablir en temps opportun le budget nécessaire aux migrations	
Suivre l'évolution des coûts (augmentation de la quantité de données)	
Entretenir l'infrastructure des archives et l'adapter à l'évolution technologique	
Documenter les données archivées (quelles données, stockées à quel endroit, etc.)	
Classement des données (type de données, accès)	
Effectuer les relectures de contrôle périodiques conformément à la planification (support, sommes de contrôle)	
Planifier la (première) prochaine migration	
Conserver les données dans des bâtiments séparés géographiquement	
Définir le lieu de la conservation (intérieur ou extérieur, en fonction des critères prix/sécurité/temps nécessaire pour y accéder)	
Budgétiser, planifier et exécuter la formation et le perfectionnement du personnel	
...	

10.2 CHAMPS PRINCIPAUX RLG (SUMMARY OF CORE FIELDS)

Saisie d'une collection (Collection record)

Données de référence (Base-line)

Collection Title (titre [de la collection])

Collection Contributor (participant(s) à la collection)

Collection Date Range (choix de la date de la collection)

Collection Narrative Description (description de la collection)

Collection Work Type (type d'ouvrage de la collection)

Collection Subjects (Topics, People, Places: (sujets de la collection [thèmes, personnes, lieux])

Saisie Objet (Individual record)

Données de référence (Base-line)

Creator (créateur)

Work Type (type d'ouvrage)

Work Title (titre de l'ouvrage)

Creation Date (date de création)

Unique ID (ID unique)

Pointer(s) to Surrogate(s) (renvoi[s] au[x] substitute[s])

Information à valeur ajoutée (Value-added)

Place of Creation (lieu de création)

Subjects (Topics, People, Places) (sujets [thèmes, personnes, lieux])

Work Description (description de l'ouvrage)

Bonus

Sequence / Hierarchy of Surrogates (Multi-part work)
(séquence / hiérarchie des éléments secondaires)

Caption for Surrogates (Multi-part work) (étiquetage des produits documentaires)

Measurement (mesure)

Materials and Techniques (matériel et méthodes techniques)

Informations détaillées: RLG (2005: 13)

10.3 VUE D'ENSEMBLE : NORMES ET SCHÉMAS DE MÉTADONNÉES

Les institutions qui se sont intéressées très tôt à la question des métadonnées – en particulier les bibliothèques – l’ont fait en raison de la nécessité de désigner leurs outils de recherche fraîchement créés (puis leurs archives numériques) par des mots-clés. Rapidement, il est apparu que les catalogues numériques ne permettaient pas seulement un accès rapide et efficace aux informations, mais offraient également des avantages par rapport aux moyens conventionnellement utilisés pour l’échange de données, ce qui était surtout intéressant pour la recherche. Les échanges étaient plus rapides et également plus économiques s’ils étaient réalisés dans les règles de l’art.

Procéder dans les règles de l’art signifie utiliser un format d’échange que l’expéditeur et le destinataire des données peuvent coder et décoder. C’est là la condition sine qua non pour qu’un échange de données réussisse, c’est également l’origine des formats d’échange (p. ex. RTF) qui conservent le formatage du texte. Dans le contexte des métadonnées, la notion d’interopérabilité est utilisée de plus en plus souvent; elle ne permet pas seulement l’échange de données, mais supporte également l’interaction entre des systèmes de conceptions différentes.

Dans l’énumération non exhaustive ci-dessous, nous présenterons brièvement les normes, schémas et initiatives les plus répandues concernant les métadonnées.

Documentation:
<http://dublincore.org>

10.3.1.1 DUBLIN CORE

Dublin Core est le schéma de métadonnées le plus répandu et est composé de 15 éléments: titre (title), créateur (creator), thème (subject), description (description), éditeur (publisher), contributeurs (contributors), date (date), type (type), format (format), identifiant (identifier), source (source), langue (language), relation (relation), champ d’application (coverage) et droits (rights).

Les éléments de données Dublin Core peuvent être définis comme des métabalises HTML ou XML, des balises RDF, ou comme attributs de champs dans des banques de données. Prévu à l’origine pour la description de documents numériques, ce schéma est également utilisé pour le classement de catalogues d’objets de musée (objets physiques).

Documentation:
www.vraweb.org

10.3.1.2 VRA (VISUAL RESOURCES ASSOCIATION) / VRA CORE

VRA est un standard pour la description d'images (œuvres visuelles) dans le domaine de la culture. Les métadonnées caractérisent l'œuvre et ses illustrations de manière hiérarchique et selon des catégories. VRA Core, qui s'appuie sur CDWA et Dublin Core, se compose de 17 éléments principaux.

Documentation:
[http://www.getty.edu/research/
conducting_research/standards/cdwa/](http://www.getty.edu/research/conducting_research/standards/cdwa/)

10.3.1.3 CDWA (CATEGORIES FOR THE DESCRIPTION OF WORKS OF ART)

Les CDWA définissent une certaine quantité d'éléments de métadonnées pour la description d'œuvres d'art et sont surtout conçues pour l'accès et la recherche. Les descriptions d'objets sont structurées selon une hiérarchie. Pour remplacer la version d'origine qui contenait au moins 500 catégories et sous-catégories, une version CDWA Lite a été créée; ses éléments principaux ont également été publiés en tant que schéma XML.

Documentation:
<http://www.knaw.nl/ECPA/sepia/>

Outil / Sepiades Tool:
[http://www.knaw.nl/ECPA/sepia/
workinggroups/wp5/download.html](http://www.knaw.nl/ECPA/sepia/workinggroups/wp5/download.html)

10.3.1.4 SEPIADES (SEPIA DATA ELEMENT SET)

SEPIADES a été mis au point en 2003 par le programme européen SEPIA (Safeguarding European Photographic Images for Access). Il est principalement dédié aux archives photographiques. Son organisation est hiérarchique et repose sur le standard de répertoire ISAD(G). Une entrée de métadonnées est d'abord créée pour une collection de photos, puis est appliquée à des groupes d'objets et, au besoin, à des objets individuels.

Documentation:
<http://www.loc.gov/marc/>

10.3.1.5 MACHINE-READABLE CATALOGING (MARC & MARCXML)

Les formats MARC21 définissent des standards pour la représentation et la transmission d'informations bibliographiques dans une forme lisible par une machine. L'échange de données se fait soit par les deux formats d'échange ISO 2709 et ANSI / NISO Z39.2, soit par XML. MARC est un «format bibliographique» pour les livres, les prises de son, les images animées et les collections d'archives; c'est la norme la plus répandue dans les bibliothèques et les archives.

Documentation:

<http://www.ifla.org/VI/3/p1996-1/unimarc.htm>

10.3.1.6 UNIMARC ET XML UNIMARC

UNIMARC et XML UNIMARC (MARC universel) sont des normes pour la représentation et la transmission d'informations bibliographiques sous une forme lisible par une machine. UNIMARC définit une base commune pour l'échange de données d'informations bibliographiques aux formats internationaux MARC.

Documentation:

www.loc.gov/standards/mods/

10.3.1.7 METADATA OBJECT DESCRIPTION SCHEMA (MODS)

Le Metadata Object Description Schema (MODS) est un schéma de métadonnées XML codé pour les informations bibliographiques. MODS contient un sous-ensemble de champs MARC21 et se positionne entre le simple Dublin Core et la description plus détaillée de MARC21. MODS est utilisé, dans la mesure du possible, dans les bibliothèques et les archives utilisant déjà MARC21 ou MARCXML. En tant que version simplifiée de ce dernier, MODS doit être davantage lisible par l'homme.

Documentation:

<http://www.mda.org.uk/index.htm>

(Etat: février 2008; actuellement lien désactivé)

10.3.1.8 SPECTRUM

SPECTRUM, The UK Museum Documentation Standard, est une norme reconnue au niveau national et international pour la documentation des musées. Il existe un schéma SPECTRUM XML, utilisé pour la description d'objets de musée.

Documentation:

http://www.english-heritage.org.uk/upload/pdf/MIDAS_Heritage_Part_One.pdf

10.3.1.9 MIDAS HERITAGE (UK HISTORIC ENVIRONMENT DATA STANDARD)

MIDAS est un standard pour les «environnements historiques» (historic environments) et établit quelles informations doivent être enregistrées pour que l'échange d'information et la conservation à long terme puissent avoir lieu de manière efficace. MIDAS respecte d'autres normes, p. ex. SPECTRUM.

Documentation:
<http://www.ica.org/fr>

10.3.1.10 INTERNATIONAL STANDARD FOR ARCHIVAL DESCRIPTION GENERAL — ISAD(G)

ISAD(G) est un standard international pour la description des fonds d'archives, utilisé avant tout dans la gestion des documents (records management). Le standard prescrit quelles informations relatives à une collection doivent être fournies, mais n'indique pas de structure explicite des données. ISAD(G) est construit par niveaux hiérarchiques (archives, stock, dossier [ensemble de documents], document).

Documentation:
<http://www.loc.gov/ead/>

10.3.1.11 ENCODED ARCHIVAL DESCRIPTION (EAD)

La description archivistique encodée «Encoded Archival Description Document Type Definition» (EAD DTD) est un standard codé SGML (Standard Generalized Markup Language, prédécesseur d'HTML) ou XML pour la description des moyens de recherche archivistique. La structure hiérarchique permet des descriptions à différents échelons et est conçue pour la combinaison d'objets numériques (collections, groupements de fonds d'archives, pièces individuelles, cf. ISAD(G)). EAD peut également être utilisée pour la description d'objets de musée.

Documentation:
<http://www.archivists.org/catalog/pubDetail.asp?objectID=1279>

10.3.1.12 DESCRIBING ARCHIVES: A CONTENT STANDARD (DACS)

DACS est une norme créée et utilisée aux Etats-Unis pour les répertoires archivistiques et se fonde sur ISAD(G). Ce standard montre comment des informations sur des ressources d'archives peuvent être intégrées dans des schémas comme EAD ou MARC21.

Documentation:
<http://www.tei-c.org/>

10.3.1.13 TEXT ENCODING INITIATIVE FOR ELECTRONIC TEXT ENCODING AND INTERCHANGE (TEI)

TEI définit un éventail d'éléments de données pour le balisage de textes en SGML ou XML. TEI-Lite est un sous-ensemble de TEI DTD et a été utilisé jusqu'ici pour la transcription électronique de livres et d'autres supports de textes. Chaque document TEI possède un en-tête (header) avec des métadonnées indiquant la source, la date de création et la transcription électronique.

Documentation:
<http://preview.tinyurl.com/3l4gf5>

Voir aussi NISO-MIX:
<http://www.loc.gov/standards/mix/>

10.3.1.14 NISO (NISO TECHNICAL METADATA FOR DIGITAL STILL IMAGES)

La norme Z39.87 de la NISO (National Information Standards Organisation) est une norme de métadonnées techniques pour les images tramées (basées sur des pixels) En même temps, cette norme sert de dictionnaire de données.

Documentation:
<http://www.loc.gov/standards/mets/>

10.3.1.15 METADATA ENCODING AND TRANSMISSION STANDARD (METS)

METS décrit des objets numériques au moyen d'un schéma XML. METS peut être utilisé pour l'échange de données, la présentation sur le Web et l'archivage numérique.

La spécification METS d'un document se compose de sept parties: l'en-tête (header), les métadonnées descriptives (descriptive metadata), les métadonnées administratives (administrative metadata), la liste des fichiers (file section), la carte de structure (structural map), les liens hypertextes entre objets (structural links), et le comportement (behavior).

En raison de la complexité des modèles XML, le codage manuel est difficile. Par conséquent, dans l'idéal les métadonnées METS sont générées à partir de banques de données.

Documentation:
<http://www.oclc.org/services/cataloging/default.htm>

10.3.1.16 PRESERVATION METADATA IMPLEMENTATION STRATEGIES (PREMIS)

PREMIS correspond moins que les autres à la définition explicite d'une norme de métadonnées. Le groupe de travail a élaboré un dictionnaire de données, qui définit une «quantité réaliste d'éléments principaux de métadonnées» pour la conservation à long terme d'objet numérique ayant des possibilités d'application étendues. Le schéma de données PREMIS contient sept entités principales: objet (object), entité intellectuelle (intellectual entity), événement (event), agent (agent), droits (right) relation (relationship) et unité sémantique (semantic unit).

Documentation:
<http://zope.cetis.ac.uk/profiles/uklomcore>

10.3.1.17 LEARNING OBJECT METADATA (LOM) ET LOM CORE

IEEE LOM est un schéma de métadonnées codé en XML pour la description de ressources didactiques. LOM Core est une version optimisée de LOM pour le domaine de l'enseignement au Royaume-Uni (UK).

Documentation:
<http://www.openarchives.org/>

10.3.1.18 OPEN ARCHIVES INITIATIVE PROTOCOL FOR METADATA HARVESTING (OAI_PMH)

OAI est un groupe de travail qui s'intéresse à la collecte (harvesting) et à la diffusion (dissemination) de métadonnées via le protocole OAI et soutient l'interopérabilité entre les différents systèmes.

Documentation:
<http://www.exif.org/>

10.3.1.19 EXCHANGEABLE IMAGE FILE FORMAT (EXIF)

Standard de métadonnées techniques pour les fichiers images, développé par l'entreprise JEITA (Japan Electronics and Information Technology Industries Association). Comme avec IPTC, les données sont écrites directement dans le fichier (TIFF et JPEG).

Documentation:
<http://www.iptc.org/>

10.3.1.20 INTERNATIONAL PRESS AND TELECOMMUNICATIONS COUNCIL (IPTC)

Schéma développé par l'IPTC pour les métadonnées descriptives, qui peuvent être stockées directement dans le fichier image (TIFF et JPEG). La description des champs IPTC est supportée par la plupart des programmes de traitement des images. Une variante basée sur XML, IPTC G2, est en cours d'élaboration.

Documentation:
<http://www.adobe.com/products/xmp/index.html>

10.3.1.21 EXTENSIBLE MARKUP LANGUAGE (XMP)

Standard de métadonnées de l'entreprise Adobe, basé sur XML et supporté par de plus en plus de logiciels fournis par des tiers. Conçu à l'origine comme standard de métadonnées pour sa propre gamme de logiciels, XMP supporte également d'autres normes et permet aussi, entre autres, l'insertion d'informations IPTC dans d'autres fichiers images que TIFF et JPEG (JPEG2000, PNG, DNG). XMP n'écrit et ne gère pas que des données IPTC. Les informations peuvent être gérées et affichées à partir de diverses applications.

10.4 VOCABULAIRES ET THÉSAURUS CONTRÔLÉS

Des vocabulaires et thésaurus contrôlés contribuent à l'uniformité et à la cohérence de la description (indexation) des objets numériques. Leur utilisation permet en outre une synonymie contrôlée, ainsi que des possibilités de recherche (filtres) et de structuration précises. Pour que cela fonctionne, il faut que l'indexation se déroule de manière méthodique. Selon le domaine, cette tâche doit être confiée à un spécialiste. En principe, les inconvénients apparaissent lors de l'élargissement obligatoire du vocabulaire.

Contrairement aux Etats-Unis, il y a peu de vocabulaires contrôlés dans le milieu germanophone. La plupart des bibliothèques utilisent le répertoire d'autorités-matières SWD.

Il est possible en tout temps de développer un vocabulaire spécifique à une application ou à un domaine. Tant que ce vocabulaire est utilisé par son «propriétaire», cela ne cause aucun problème. Par contre, dès que des données doivent être échangées, l'interopérabilité avec d'autres systèmes ne peut être garantie que moyennant un investissement en temps et des coûts supplémentaires. C'est la raison pour laquelle les institutions travaillant les unes avec les autres devraient utiliser un vocabulaire semblable, sauf exception justifiée.

Dans les pays anglo-saxons, il existe un grand nombre de vocabulaires contrôlés, qui ne sont pas d'une grande pertinence ni pour les régions linguistiques germanophones, ni pour la Suisse. Cependant, les sources ci-dessous fournissent des informations utiles:

www.getty.edu/research/conducting_research/vocabularies/aat/

10.4.1.1 ART & ARCHITECTURE THESAURUS® (AAT)

Vocabulaire structuré/terminologie relatifs à l'art, à l'architecture et à des disciplines appliquées.

www.getty.edu/research/conducting_research/standards/cdwa/

10.4.1.2 CATEGORIES FOR THE DESCRIPTION OF WORKS OF ART (CDWA)

Plutôt un schéma de métadonnées, qui décrit des catégories pour l'art, l'architecture, etc. (cf. 10.3.1.3).

www.getty.edu/research/conducting_research/vocabularies/tgn/

10.4.1.3 GETTY THESAURUS OF GEOGRAPHIC NAMES (TGN)

Vocabulaire structuré, en particulier pour les descriptions géographiques.

www.loc.gov/rr/print/gm/graphmat.html

10.4.1.4 GRAPHIC MATERIALS

Règles pour la création de catalogues d'impressions, de photographies et d'autres images.

www.ica.org/fr

10.4.1.5 INTERNATIONAL STANDARD ARCHIVAL AUTHORITY RECORD FOR CORPORATE BODIES, PERSONS, AND FAMILIES (ISAAR (CPF))

Contient 26 éléments assurant l'identification correcte de matériel propre aux archives.

www.iconclass.nl

10.4.1.6 ICONCLASS

Système de classification pour la recherche iconographique, la documentation en images, etc.

<http://authorities.loc.gov/>

10.4.1.7 LIBRARY OF CONGRESS SUBJECT HEADINGS (LCSH)

Le plus grand vocabulaire général d'indexation en anglais.

<http://www.loc.gov/catdir/cpsd/amimupd.html>

10.4.1.8 MOVING IMAGE MATERIALS: GENRE TERMS

Recueil terminologique pour l'indexation de films.

<http://www.rarebookstore.net/cgi-bin/schuyler/5577.html?id=jq6air9t>

10.4.1.9 SUBJECT INDEX FOR THE VISUAL ARTS

Vocabulaire structuré pour l'accès aux objets de la Print Room du Victoria & Albert Museum de Londres.

www.loc.gov/rr/print/tgm1/

10.4.1.10 THESAURUS FOR GRAPHIC MATERIALS (TGM)

www.loc.gov/rr/print/tgm2/

Vocabulaire structuré, contenant des mots-clés pour le matériel visuel.

www.getty.edu/research/conducting_research/vocabularies/ulan/

10.4.1.11 UNION LIST OF ARTIST NAMES® (ULAN)

Vocabulaire structuré contenant les noms et d'autres informations concernant les créateurs d'art.

11 LISTE DES ABRÉVIATIONS

ANSI	American National Standards Institute
CDWA	Categories for the Description of Works of Art
Cedars	Curl (Consortium of Research Libraries) exemplars in digital archives
DACS	Describing Archives a Content Standard
DCMI	Dublin Core Metadata Initiative
DNG	Digital Negative (Adobe)
EAD	Encoded Archival Description (description archivistique encodée)
Exif	Exchangeable Image File Format (format d'échange de fichiers images)
HTML	Hypertext Markup Language (langage hypertexte)
IEEE	Institute of Electrical and Electronics Engineers
IPTC	International Press Telecommunication Council
IPTC4XMP	IPTC Metadata for XMP
ISAD(G)	International Standard for Archival Description (General)
JPEG	Joint Photographic Experts Group (Groupe mixte d'experts en photographie)
JPEG2000	Nouveau format de fichier du Groupe mixte d'experts en photographie
JPG	Extension courante des fichiers JPEG
LOM	Learning Object Metadata
MARC	Machine Readable Cataloging (record)
MARC21	Famille de formats MARC Concise
MARXML	MARC en tant que schéma XML
MD5	Message-Digest Algorithm
METS	Metadata Encoding and Transmission Standard
MIDAS	UK Historic Environment Data Standard
MODS	Metadata Object Description Schema
NISO	National Information Standards Organisation
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAIS	Open Archival Information System
OCLC	Online Computer Library Center
PADI	Preserving Access to Digital Information
PDF	Portable Document Format (Adobe) (format de document portable)
PNG	Portable Network Graphics

PREMIS	PREservation Metadata Implementation Strategies
PSD	Photoshop Data Format (format de données Photoshop)
RDF	Resource Description Framework (canevas de description de ressource)
RLG	Research Libraries Group
SEPIA	Safeguarding European Photographic Images for Access
SEPIADES	SEPIA Data Element Set
SGML	Standard Generalized Markup Language (langage standard de balisage généralisé)
SPECTRUM	The UK Museum Documentation Standard
TASI	Technical Advisory Service for Images
TEI	Text Encoding Initiative
TIFF	Tagged Image File Format (fichier d'image étiqueté)
UNIMARC	Format MARC pour l'échange international
VRA Core	Visual Resources Association Core
XML	eXtensible Markup Language (langage de balisage extensible)
XMP	Extensible Metadata Platform

NOTES

ADRESSES DE CONTACT

MANDANT

Office fédéral de la protection de la population OFPP
Protection des biens culturels PBC
Rino Büchel, Reto Suter
Monbijoustrasse 51A
3003 Berne
www.kulturgueterschutz.ch → Français
Tél.: +41 (0)31 325 15 37
Fax: +41 (0)31 324 87 89
E-Mail: reto.suter@babs.admin.ch

ADRESSE DE CONTACT POUR TOUTE INFORMATION CONCERNANT LES MÉTADONNÉES

Memoriav
Verein zur Erhaltung des
audiovisuellen Kulturgutes der Schweiz
Effingerstrasse 92
3008 Berne
www.memoriav.ch
Tél.: +41 (0)31 380 10 80
Fax: +41 (0)31 380 10 81
E-Mail: info@memoriav.ch

AUTEURS DES GUIDELINES

Imaging and Media Lab
Université de Bâle
Sergio Gregorio, Anja-Elena Stepanovic
Bernoullistrasse 32
4056 Bâle
www.iml.unibas.ch
Tél.: +41 (0)61 267 38 36
Fax: +41 (0)61 267 04 85
E-Mail: iml@unibas.ch