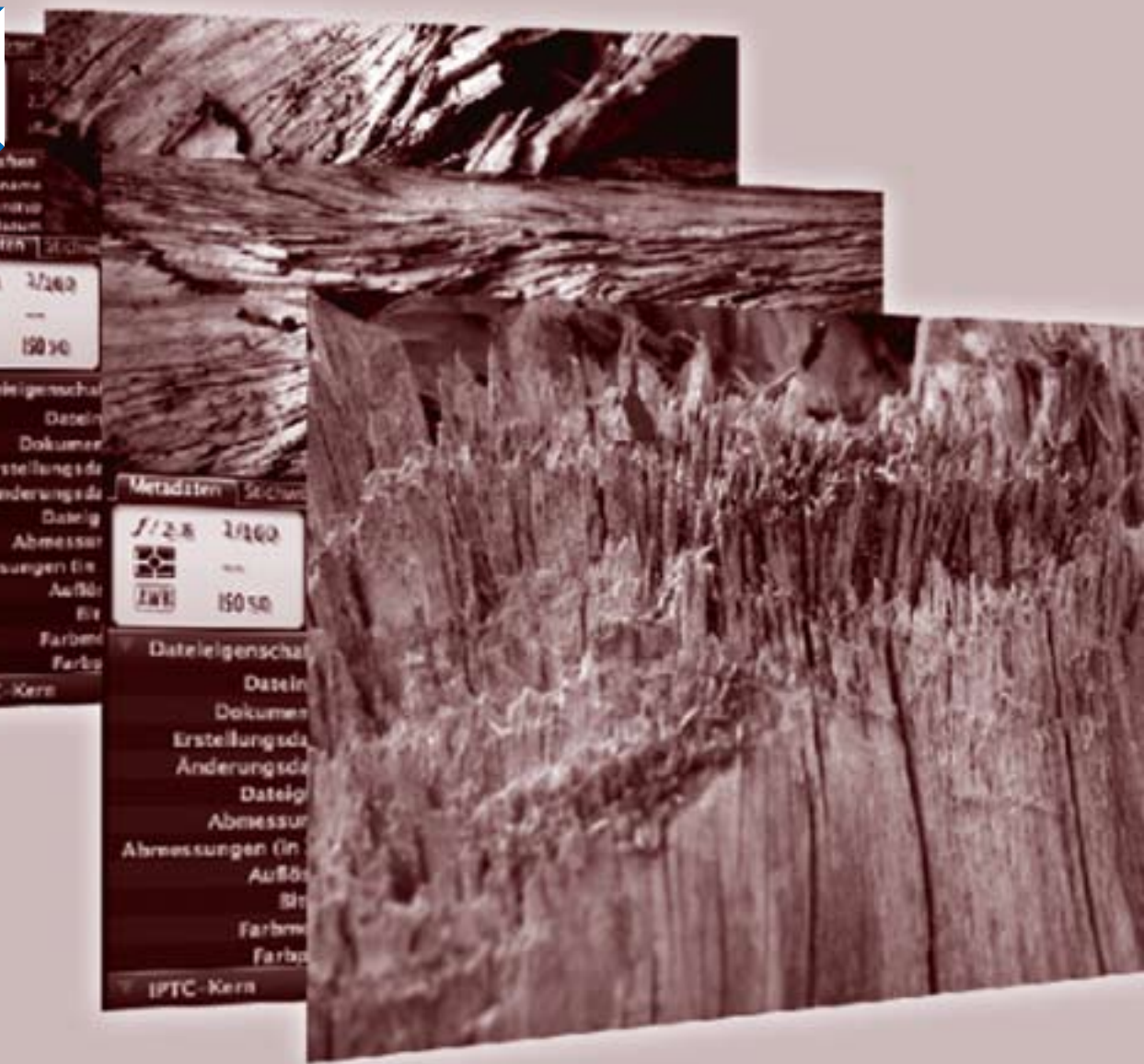


> 03.2008

> THEMA: METADATEN BEI STEHENDEN* DIGITALEN BILDERN
> AUTOREN: SERGIO GREGORIO, ANJA-ELENA STEPANOVIC

GUIDELINES



METADATEN BEI STEHENDEN* DIGITALEN BILDERN

* FÜR DIE PRÄZISIERUNG «STEHENDE DIGITALE BILDER (DIGITAL STILL IMAGES)» GILT IN DIESEN GUIDELINES IN DER FOLGE ÜBERALL DIE ELLIPTISCHE VERSION DIGITALE BILDER.

TITELBILD

Illustration gängiger Bild-Dateiformate.
Bild: Elias Kreyenbühl, Imaging and Media Lab



Mix

Produktgruppe aus vorbildlich bewirtschafteten
Wäldern und anderen kontrollierten Herkünften
www.fsc.org Zert.-Nr. SQS-COC-100005
© 1996 Forest Stewardship Council

METADATEN BEI STEHENDEN*

DIGITALEN BILDERN

*Für die Präzisierung stehende «digitale Bilder (digital still images)» gilt in diesen Guidelines in der Folge überall die elliptische Version digitale Bilder.

AUTOREN:

SERGIO GREGORIO & ANJA-ELENA STEPANOVIC

Stand: 28.02.2008

AUFTRAGSERTEILUNG, PROJEKTBEGLEITUNG:

RINO BÜCHEL, RETO SUTER

Auftraggeber:

© Bundesamt für Bevölkerungsschutz, Fachbereich Kulturgüterschutz, Bern, 2008

Redaktion:

Reto Suter, Hans Schüpbach

Layout:

Zentrum elektronische Medien ZEM, Bern

VORWORT

*Rino Büchel
Chef Kulturgüterschutz (KGS) im
Bundesamt für Bevölkerungsschutz
BABS, Bern.*

*Prof. Dr. Rudolf Gschwind
Imaging and Media Lab, Basel.*

Metadaten werden in der Regel vereinfacht als «Daten über Daten» paraphrasiert. Ohne praktischen Bezug sagt diese kurze Charakterisierung jedoch wenig aus. Im digitalen Kontext beschreiben Metadaten digitale Objekte (Text-, Bild- und Tondateien), erleichtern deren Auffindbarkeit und unterstützen die Informationsvermittlung. Obwohl ausführliche Dokumentationen zu verschiedenen Aspekten im Internet erhältlich sind, bleibt das Thema Metadaten selbst für den interessierten Benutzer abstrakt, diffus und manchmal schwer fassbar. Ein Beispiel aus der physischen Welt soll den Sachverhalt etwas besser veranschaulichen.

Szenario: Nach einer Hochwasserkatastrophe werden Keller ausgepumpt, gereinigt und aufgeräumt. Vorräte, die nicht mehr zu retten sind, werden entsorgt. Was aber soll mit dem ungeordneten Haufen Dosen geschehen, die durch das Hochwasser ihre Beschriftung verloren haben (Abb. 2)? Der Inhalt hat die Katastrophe unbeschadet überstanden, ist aber von aussen und ohne vorheriges Öffnen der Dosen nicht mehr nachvollziehbar. Wie werden z. B. Ravioli von Hundefutter unterschieden?

Im digitalen Kontext verhält es sich ähnlich. Ein Dokument, in unserem Fall ein digitales Bild, das keine Beschreibung hat – also keine Metadaten aufweist –, muss zuerst geöffnet werden, damit erkennbar wird, worum es sich handelt.



Abb. 1 und 2 veranschaulichen an einem Beispiel die Problematik fehlender Metadaten.

1 Dosen mit Aufschrift (= mit Metadaten).

2 Dosen ohne Aufschrift (= fehlende Metadaten).

Es gibt aber auch Dosen mit aufgedruckten Beschreibungen (vgl. Dose ganz rechts im Bild 2), bei denen auch nach dem Wasserschaden noch ersichtlich ist, was sie enthalten. Diese sind durch den Mehraufwand des Aufdrucks teurer. In Analogie dazu sind in der digitalen Welt die Metadaten im Objekt integriert, was zwar mit Mehraufwand verbunden ist und damit höhere Produktionskosten zur Folge hat. Dafür können Papieretiketten in diesem Fall als extern gespeicherte Metadaten (meistens in einer separaten Datei oder in einer Datenbank) betrachtet werden, die höchstwahrscheinlich verloren gehen, falls sie vom Objekt getrennt werden. Der Vergleich mag nicht allen Detailbetrachtungen standhalten. Wichtig ist, dass digitale Objekte vor allem im Hinblick auf eine dauerhafte Archivierung mit Metadaten versehen werden, damit diese nicht nur auffindbar, sondern auch in Zukunft interpretierbar bleiben.

Genauso wie in der physischen Welt gehören auch im digitalen Kontext die Metadaten (die Beschriftung, das Etikett) zum Objekt und sollten, wann immer möglich, als Bestandteil in die entsprechende Datei integriert werden.

Im Bereich der digitalen Bilder ist vieles noch offen und der Umgang mit Metadaten noch nicht zufriedenstellend gelöst. Diese Tatsache darf aber nicht dazu verleiten, digitale Bildbestände nicht gepflegt auf einen Datenträger auszulagern und «auf bessere Zeiten» zu warten. Bei diesem Vorgehen nehmen die Wahrscheinlichkeit und das Risiko eines Datenverlusts mit dem Fortschreiten der Zeit stetig zu. Nur eine aktive Bewirtschaftung der digitalen Bildbestände garantiert aus heutiger Sicht das Fortbestehen von digitalem Kulturgut.

Die vorliegenden Guidelines führen in das Thema Metadaten bei digitalen Bildern ein, dokumentieren eine erprobte Vorgehensweise, wie Metadaten definiert werden können, und zeigen weitere Schritte auf, die ein digitaler Bildbestand im Hinblick auf eine dauerhafte Archivierung durchlaufen sollte. Insbesondere sei auf die Checkliste (Kap. 10.1) verwiesen, die den Verantwortlichen von Institutionen als nützliche Orientierungshilfe dienen kann.

Mehrere Personen, die nachstehend in alphabetischer Reihenfolge genannt werden, haben Kommentare zum Entwurf der Guidelines abgegeben – ihre Bemerkungen wurden in die Erarbeitung der Endfassung einbezogen. Für die kritische Durchsicht des Textes bedanken wir uns herzlich bei Kurt Deggeller (Direktor Memoriav), Urs Peter Schelbert (Staatsarchivar Kt. Zug), Bernard A. Schüle (Leiter Objektzentrum der Schweizerischen Landesmuseen), Daniel Stadlin (KGS-Verantwortlicher Kt. Zug) und Dr. Tobias Wildi (Docuteam GmbH).

Dieser Leitfaden soll kleine und mittlere kulturelle Institutionen (Archive, Bibliotheken, Museen und verwandte Institutionen) bei der Bewirtschaftung und Archivierung ihrer digitalen Bildbestände unterstützen. In diesem Sinne möchte der Fachbereich Kulturgüter-schutz (KGS BABS) in Zusammenarbeit mit dem Imaging and Media Lab der Universität Basel einen grundlegenden Beitrag gegen das «digitale Vergessen» leisten.

INHALTSVERZEICHNIS

1	Einführung	8
2	Bild-Dateiformate	9
2.1	TIFF	9
2.2	JPEG	10
2.3	JPEG2000	10
2.4	Fazit	11
3	Metadaten	12
3.1	Klassifikation	13
3.2	Fazit	14
4	Metadatenstandards	15
4.1	Fazit Bild-Metadaten	17
5	Anwendungsbeispiel	18
6	Kernfelder	23
7	Digitale Archivierung	24
7.1	Archivierungsprozess	25
7.1.1.	Vorbereiten	25
7.1.2.	Bearbeiten	25
7.1.3.	Kopieren	26
7.1.4.	Bewirtschaften	27
7.2	Fazit	27
8	Schlussfolgerungen und Ausblick	28
9	Bibliografie	30
10	Anhang	32
10.1	Checkliste	32
10.2	RLG Kernfelder	34
10.3	Übersicht: Metadatenstandards und -schemata	35
10.3.1.1–10.3.1.21	Einzelne Standards und Schemata	35
10.4	Kontrollierte Vokabulare und Thesauri	41
10.4.1.1–10.4.1.11	Einzelne Vokabulare und Thesauri	41
11	Abkürzungsverzeichnis	44

1 EINFÜHRUNG

Rapide anwachsende Datenbestände und die stetig zunehmende Flut an ausschliesslich digital verfügbaren Dokumenten erschweren die Überschaubarkeit und die gezielte Suche nach Informationen. Oft sind Suchergebnisse trotz Einsatz modernster Findmittel unvollständig, da ein digitales Objekt ohne Beschreibung schwer zu finden ist. Diese Tatsache trifft insbesondere auf digitale Bilder zu, deren Inhalt nicht anhand einer einfachen Textsuche nachträglich erschlossen werden kann. Deshalb sind Bildbeschreibungen für digitale Bilder von fundamentaler Bedeutung. Solche beschreibende Metadaten sind auch für den Kontexterhalt archivierter Objekte eine unabdingbare Voraussetzung und erlauben als untrennbarer Bestandteil eines digitalen Objekts (Text, Bild, Bewegtbild, Ton) nicht nur das Auffinden, sondern auch ein zukünftiges Verstehen des historischen Kontextes.

Während das Auffinden von Informationen durch eine sinnvolle Beschreibung der Objekte verbessert werden kann, wird dem Aspekt der digitalen Archivierung in der Regel weniger Beachtung geschenkt. Die vergleichsweise einfache Nutzung von Informationstechnologie täuscht über die langfristige «Haltbarkeit» digitaler Objekte hinweg. Solange Zugriff und Interpretierbarkeit gegeben sind, wird auch der Langzeiterhalt digitaler Daten vorausgesetzt. Ohne eine langfristig ausgerichtete digitale Archivierstrategie ist ein Datenverlust – und der damit verbundene Informationsverlust – unausweichlich. Deshalb muss dieser Aspekt im Umgang mit digitalen Daten unbedingt berücksichtigt werden.

Die vorliegenden Guidelines führen in das Thema «Metadaten bei digitalen Bildern» ein. Sie zeigen auf, wie Bilddateien mit Informationen versehen und damit auch für den Langzeiterhalt vorbereitet werden können. Die einzelnen Schritte werden anhand eines Archivierungsprozesses veranschaulicht. Einführend werden die heute wichtigsten Bilddateiformate vorgestellt und die verschiedenen Ausprägungen von Metadaten erklärt. Es folgt eine Übersicht über aktuelle Metadatenstandards, unter denen es zurzeit keinen De-iure-Standard für digitale Bilder gibt. Analoge Verfahren wie die Sicherstellung auf Mikrofilm werden nicht erwähnt und auf Aspekte der Schriftgutverwaltung wird nur marginal eingegangen. Die nachfolgenden Ausführungen sollen insbesondere kleine und mittlere Institutionen aus dem Kulturbereich (Archive, Bibliotheken, Museen und verwandte Institutionen) bei der Verwaltung, Bewirtschaftung und Archivierung ihrer Bildbestände unterstützen.

2 BILD-DATEIFORMATE

¹ Alle im Dokument verwendeten, nicht näher erklärten Abkürzungen sind im Abkürzungsverzeichnis auf S. 44 aufgeführt.

² Eine ausführliche Übersicht über aktuelle und obsolekte Bild-Dateiformate findet man unter: http://www.lemkesoft.de/xd/public/content/index._cGlkPTY5_.html

³ Ein proprietärer Standard gehört einer Firma, welche die Spezifikation gar nicht oder nur zum Teil offenlegt (z. B. das aktuelle Word-Dokumentformat von Microsoft: Stand Word 2003).

⁴ Die übliche Erweiterung ist «PSD». Das Photoshop-Format unterstützt unter anderem Ebenen und Farbprofile.

⁵ TIFF besteht aus «Feldern» mit einer Beschreibung (Tag oder Etikett), die einen Wert beinhalten: z. B. Bildgröße, Auflösung usw.

⁶ In Zeilen und Spalten angeordnete Bildpunkte (Pixel).

⁷ Das TIFF-Format wurde 1992 von der Firma Aldus Corporation entwickelt, die mit Aldus Pagemaker zu den ersten Anbietern von Textsatzsoftware zählte. 1994 erfolgte der Zusammenschluss mit Adobe.

⁸ Die TIFF-Spezifikation kann man auf der Adobe Website unter folgender Adresse herunterladen:
<http://partners.adobe.com/asn/developer/PDFS/TN/TIFF6.pdf>
TIFF kann mit privaten Tags (private tags) erweitert werden.

Ein Dateiformat ist eine genau definierte Folge von Bits (0 und 1), die als Bitstrom bezeichnet wird. Diese Folge ist genau definiert, damit der Computer sie als eine Text-, Bild- oder Tondatei erkennen und entsprechend verarbeiten kann. Für Dateiformate existieren Spezifikationen, die alle charakteristischen Eigenschaften dokumentieren.

Für digitale Bilder existieren unzählige Dateiformate, von denen sich für die Nutzung und für den Langzeiterhalt im Prinzip nur JPEG¹ und TIFF durchgesetzt haben.² Nicht nur für die Nutzung, sondern auch für die Archivierung ist wichtig, dass es sich um offene Formate handelt, die auf einem offenen Standard³ beruhen. Ein offener Standard ist im Vergleich zu einem proprietären Standard ausführlich dokumentiert und im Detail nachvollziehbar.

Wenn JPEG und TIFF für die Nutzung und für den langfristigen Erhalt eingesetzt werden, befindet man sich aus heutiger Sicht und Praxis auf der sicheren Seite. Natürlich sind andere Formate wie z. B. jenes von Adobe Photoshop⁴ für die Bearbeitung erlaubt. Sobald diese abgeschlossen ist, empfiehlt es sich jedoch, das Ergebnis in ein für die Archivierung geeignetes Format zu konvertieren, weil sich proprietäre Formate von kommerziellen Softwareanbietern durch die immer rasanteren Innovationszyklen rascher ändern und damit auch schneller unlesbar werden. Beispielsweise sollten von einem Bild nach Abschluss aller notwendigen Bildbearbeitungsschritte eine oder mehrere JPEG-Versionen für die Nutzung sowie eine TIFF-Version für die Archivierung erstellt werden.

2.1 TIFF

TIFF ist das Akronym für «Tagged Image File Format» und ein Format, das auf Auszeichnungen oder Tags⁵ aufbaut. Es wurde für die Verarbeitung von Bitmuster-Bildern⁶ und Grafiken konzipiert.

TIFF – die übliche Dateierweiterung ist .TIF – existiert seit 1992⁷ und ist seither weiterentwickelt worden. Mit neuen Anforderungen kamen neue Tags hinzu. TIFF umfasst deshalb mehrere Standards.⁸

Die Vorteile von TIFF sind die weite Verbreitung, eine gewisse Robustheit im Sinne einer vergleichsweise geringeren Fehleranfälligkeit und die Tatsache, dass es sich um ein sogenanntes unkomprimiertes Format handelt. Für den Benutzer bedeutet dies konkret, dass z. B. mehrmaliges Speichern keinen Informationsverlust zur Folge hat, d. h. eine TIFF-Datei ist beliebig veränderbar, ohne dass dabei Bildinformationen verloren gehen, was aber nichts über den visuellen Bildgehalt aussagt.

Nachteile von TIFF sind die Komplexität und vor allem die Dateigröße, die bei grösseren Datenmengen eine erhebliche finanzielle Belastung zur Folge hat. Für die Langzeitarchivierung digitaler Bilder ist TIFF weiterhin das Dateiformat der Wahl, obwohl aus ökonomischer Sicht die Datenkompression vermehrt in den Diskurs über die dauerhafte Archivierung von digitalen Daten rückt.

⁹JPEG wurde 1990 standardisiert und danach zusehends von Softwareanbietern unterstützt und genutzt. Die Spezifikation ist unter folgender Adresse erhältlich: <http://www.w3.org/Graphics/JPEG/jfif3.pdf>

¹⁰ IPTC wird in Kapitel 3 beschrieben.

¹¹ Sollen vor der Archivierung JPEG-Bilder bearbeitet werden, empfiehlt es sich, diese zuerst in TIFF zu konvertieren und nach Abschluss der Bearbeitung, falls nötig, daraus wieder JPEG-Versionen zu erstellen. Somit kann sichergestellt werden, dass durch die verlustbehaftete JPEG-Kompression kein Datenverlust eintritt.

¹² Für die geringe Verbreitung von JPEG2000 ist neben der zum Format gehörenden Komplexität möglicherweise auch die mangelnde Unterstützung durch die bekanntesten Web-Browser verantwortlich. Eine Aufzählung bekannter Softwarepakete, die JPEG2000 unterstützen, findet sich unter: http://de.wikipedia.org/wiki/JPEG_2000

¹³ Die Erläuterung dieser primär technischen Aspekte würde den Rahmen der vorliegenden Publikation sprengen. Wikipedia liefert eine gute Übersicht und Erklärungen zu diesen Themen: z. B. unter <http://de.wikipedia.org/wiki/Wavelet-Transformation>

2.2 JPEG

JPEG ist das Akronym für «Joint Photographic Experts Group» und ein Format, das in letzter Zeit vor allem durch den Boom der Consumer-Digitalkameras eine sehr grosse Verbreitung gefunden hat.⁹ JPEG – die übliche Dateierweiterung ist .JPG – ist ein komprimierendes und verlustbehaftetes Format, das eine im Vergleich zu TIFF geringe Dateigrösse mit einem ansprechenden visuellen Erscheinungsbild vereint. Für die Benutzer bedeutet dies, dass jedes erneute Speichern einer JPEG-Datei einen Informationsverlust zur Folge hat. Dieser ist durch eine sich verringernde Dateigrösse und durch zunehmende Bildartefakte in Form von deutlich sichtbaren und störenden Mosaiken wahrnehmbar. Ursprünglich ging es darum, ein visuell ansprechendes und für die Datenübertragung vergleichsweise schlankes Bildformat zu nutzen.

Die Vorteile von JPEG sind die weite Verbreitung – und damit eine implizite längerfristige Lesbarkeit – sowie die einfache Erfassung von IPTC-Metadaten.¹⁰ Zu den Nachteilen gehören die verlustbehaftete Kompression und eine im Vergleich zu TIFF viel geringere Robustheit, d. h. Fehler innerhalb der Datei sind sofort sichtbar.

JPEG wird als verlustbehaftetes Format grundsätzlich nicht für eine dauerhafte digitale Archivierung empfohlen. Allerdings wird aus ökonomischer Sicht auch für digitale Langzeitarchive vermehrt die Datenkompression in Erwägung gezogen.¹¹

2.3 JPEG2000

Seit einigen Jahren wird das neue Bild-Dateiformat JPEG2000, eine Weiterentwicklung von JPEG, als zukünftiges Archiv-Dateiformat und als Alternative zu TIFF angepriesen. Bei näherer Betrachtung der Eigenschaften dieses neuen Dateiformats sind Erwartungen und Hoffnungen durchaus berechtigt. Das Format ist noch nicht weit verbreitet, weil noch wenige Softwareprogramme JPEG2000 vollumfänglich unterstützen.¹²

JPEG2000 ist ein vollkommen neues Bild-Dateiformat, das mit JPEG und mit TIFF nicht mehr vergleichbar ist. Es basiert auf der sogenannten diskreten Wavelet-Transformation, die Bilder schrittweise aufbaut und fließend verlustbehaftete sowie verlustfreie Kompression anwenden kann.¹³ Vorteile von JPEG2000 sind weniger wahrnehmbare Block-Artefakte (störende Mosaiken) im visuellen Bereich und eine grosse Robustheit (geringe Fehleranfälligkeit) auf der Formatenebene, was für die dauerhafte Archivierung von Bedeutung ist.

Während die direkte Erstellung von JPEG2000-Bildern keine Probleme verursachen sollte, ist bei einer zukünftigen Formatmigration darauf zu achten, dass weder Datenverlust noch Qualitätseinbussen auftreten und die Migration reversibel ist.

2.4 FAZIT

Mit der Archivierung von Bildern im TIFF-Format ist man weiterhin auf der sicheren Seite. Doch heutzutage liegen digitale Bilder, die direkt ab Digitalkamera übertragen werden, fast ausschliesslich im JPEG-Format vor. Der Aufwand, diese vor der Archivierung in TIFF umzuwandeln, kann insbesondere bei grossen Bildbeständen sehr hoch sein. Auch wenn der orthodoxe Ansatz ein archivtaugliches Format wie TIFF vorschreibt, kann die Archivierung von JPEG-Dateien in der Praxis durchaus sinnvoll sein, wenn die Vorbereitungsarbeiten korrekt durchgeführt werden und eine spätere Formatmigration, z. B. nach JPEG2000, rechtzeitig und nach allen nötigen Vorkehrungen erfolgt.

Eine Formatmigration oder -konversion wird nötig, wenn das Format obsolet wird, d. h. wenn neue Software das Format nicht mehr unterstützt oder wenn das neue gegenüber dem älteren Format klare Vorteile aufweist. Weitere wichtige Punkte, die es speziell für den Bereich der digitalen Bilder zu beachten gilt, sind in Rosenthaler (2007) dokumentiert.

3 METADATEN

¹⁴ «*Metadata is structured information that describes, explains, locates, or otherwise makes it easier to manage an information resource.*»

¹⁵ «*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>

¹⁶ «*[...] any formal scheme of resource description, applying to any type of object, digital or non-digital.*»

¹⁷ *Weitere Aspekte bezüglich Primärdaten (Datenobjekte) und Sekundärdaten (Metadaten) werden in Margulies (2007) genauer differenziert.*

¹⁸ *Der ISO-Standard 15489, Information und Dokumentation, Schriftgutverwaltung, besteht aus zwei Teilen, einem allgemeinen Teil ISO 15489-1 (Nutzen, Regelungsumfeld, Grundsätze, Zielvorgaben, Verantwortlichkeiten, Anforderungen, Schriftgutverwaltungssystem, Prozesse und Steuerung, Prüfung, Aus- und Fortbildung) und einem spezifischen Teil, Richtlinien, ISO 15489-2 (Anwendungsbereich, Grundsätze und Zuständigkeiten, Strategie, Entwicklung, Umsetzung, Überwachung und Ausbildung). ISO 23081 beschreibt Metadaten für Verfahren der Schriftgutverwaltung (Verständlichkeit der Unterlagen, Zugänglichkeit, Benutzbarkeit, Schutz, langfristige Wiederherstellung und Migration, Interoperabilität). Eine vortreffliche Zusammenfassung dieser beiden ISO-Standards liefert Siegrist (2007).*

¹⁹ *Der ISO-Standard 15836:2003 definiert z. B. nur die Dublin Core-Elemente (dublin core element set), die genutzt werden können.*

²⁰ *Z. B. zwischen ISAD(G), EAD, MARCXML, PREMIS usw. Die MIT-Libraries liefern eine Übersicht über gängige Metadaten-Mappings:*

Was sind Metadaten? Sieht man von der Bezeichnung «Daten über Daten» ab, werden Metadaten beispielsweise genauer definiert als:

Strukturierte Informationen, die eine Informationsressource beschreiben, erklären, finden oder anderenfalls die Verwaltung vereinfachen (NISO 2004),¹⁴ wobei eine Ressource eine Publikation, ein Bericht, eine Zeitschrift, eine Website über ein bestimmtes Thema, über eine Dienstleistung oder ein digitales Objekt wie z. B. ein Bild sein kann (National Library of Australia).¹⁵

Im Bibliothekswesen finden Metadaten im Allgemeinen ihre Anwendung als:

Formales Schema für die Beschreibung einer Ressource, das auf jeden beliebigen digitalen oder analogen Objekttypen anwendbar ist (NISO 2004: 1).¹⁶

Diese Definitionen dokumentieren einerseits den beschreibenden, andererseits den technischen Charakter von Metadaten, die Informationen über andere Daten oder Primärdaten enthalten. Diese Beschreibungen beinhalten Informationselemente, die den Primärdaten in der Regel nicht unmittelbar zu entnehmen sind.¹⁷ Das Thema Metadaten ist einerseits weiterhin Gegenstand der Forschung und deshalb als «work in progress» zu betrachten. Andererseits werden Standards und Schemata unentwegt weiterentwickelt. Zurzeit fehlt jedoch ein De-iure-Standard für digitale Bilder wie dies für die Schriftgutverwaltung (Records Management)¹⁸ der Fall ist.¹⁹ Bei diesen Standards und Schemata handelt es sich um Beschreibungen und Richtlinien, die im Anwendungsfall noch explizit umzusetzen sind. Die konkrete technische Implementation, auf die in der Regel nicht eingegangen wird, liegt in der Verantwortung der jeweiligen Institution.

Für die Charakterisierung digitaler Objekte stehen mehrere Metadatenschemata zur Verfügung, die als De-facto-Standards (z. B. Dublin Core, EAD, MARCXML) alle zum Ziel führen. Der Informationsaustausch ist nicht immer einfach. Die oft zitierte Interoperabilität ist mit grossem Aufwand verbunden und nur mit soliden technischen Kenntnissen zu erreichen, wenn auch im Bereich der Schriftgutverwaltung zwischen einigen Standards sogenannte Crosswalks definiert sind.²⁰ Eine Übersicht über verbreitete Standards befindet sich im Anhang 10.3.

<http://libraries.mit.edu/guides/subjects/metadata/mappings.html>

Harpring et al. (2006) liefern eine umfassende Übersicht mit Beispielen für die bekanntesten Standards.

²¹ Die unterschiedlichen Ausprägungen werden in der Regel nach ihrer Funktionalität klassifiziert. Abgrenzungen sind nicht immer eindeutig. Die ISBN-Nummer eines Buches kann z. B. einer beschreibenden oder einer administrativen Kategorie zugewiesen werden.

²² Diese werden oft als «resource discovery metadata» umschrieben.

²³ Man spricht deshalb auch von «eingebetteten Metadaten» (embedded metadata) im Gegensatz zu «stand-alone» Metadaten (nicht eingebettete, allein stehende Metadaten).

3.1 KLASSIFIKATION

Metadaten werden ihrer Funktion entsprechend in Kategorien eingeteilt. Eine weit verbreitete Kategorisierung beschränkt sich auf drei Kategorien: beschreibende, strukturelle und administrative Metadaten (Hurley et al. 1999).²¹ Ein anderes Modell sieht fünf Kategorien vor und zwar für Beschreibung (resource description), Suche (information retrieval), Verwaltung (management), Besitz (ownership and authenticity) und Interoperabilität (interoperability) (Haynes 2004: 15–17). Eine grundsätzliche Kritik liegt darin, dass in den meisten Klassifizierungen dem Nutzungskontext zu wenig Beachtung zukommt (Day 2005: 21). Auch werden technische Metadaten in diesen beiden Modellen nicht explizit erwähnt. Diese gehören genauso zum digitalen Objekt, auch wenn ihre Definition in der Regel automatisch und im Hintergrund erfolgt. In den nachfolgenden Absätzen werden beschreibende, strukturelle, administrative und technische Metadaten im Kontext von Bild-Dateien näher beschrieben. Auf diese vier Kategorien wird oft verwiesen. Es ist deshalb davon auszugehen, dass ein Konsens besteht und diese Kategorisierung längerfristig Bestand hat.

Beschreibende Metadaten (descriptive metadata) dienen der Identifikation und der Suche digitaler, aber auch analoger Objekte.²² Zu ihnen zählen unter anderem Elemente wie Titel, Autor, Thema, Schlagwort sowie weitere Beschreibungen wie z. B. physische Eigenschaften eines Objekts (Grösse, Medium und Zustand). Oft werden beschreibende Metadaten extern verwaltet, z. B. in Datenbanken oder in Textdateien, die zusammen mit dem Objekt abgelegt werden. Bei digitalen Bildern sollten diese Informationen zusätzlich in die Bild-Datei «eingebettet» (eingefügt) werden und nicht in einer zweiten Datei «für sich alleine stehen».²³ Die Verbindung zwischen Daten und Metadaten bleibt erhalten und es wird auf einfache Weise dafür gesorgt, dass Metadaten und Objekt zusammen aktualisiert werden (NISO 2004). Deshalb wird empfohlen, bei digitalen Bildern den Objekt-Identifikator (Objekt-ID: identifiziert das Objekt eindeutig), Titel oder Überschrift, Ersteller und Verleger in die Datei zu integrieren (Puglia et al. 2004).

Strukturelle Metadaten beschreiben die Beziehungen zwischen den verschiedenen Komponenten einer digitalen Ressource, fassen die einzelnen Objekte zu einer Einheit zusammen und zeigen auf, wie diese gegliedert ist (Inhaltsverzeichnisse, Kapitel usw.). Strukturelle Metadaten beschreiben zudem auch den Aufbau komplexer Objekte (zusammengesetzte Objekte), die aus mehreren und oft unterschiedlichen Einzelobjekten bestehen.

Diese Informationen werden z. B. bei der Nutzung digitaler Ressourcen für die Anzeige und für die Navigation zwischen den einzelnen Objekten eingesetzt.

²⁴ Ein Schlagwortkatalog verweist beispielsweise auf den physischen Standort eines Objekts. Nicht selten liegen zwischen der Information über das Objekt und dem Objekt selbst grössere geografische Distanzen. Die unmittelbare Nutzung wird dadurch erschwert.

Administrative Metadaten dienen vor allem dem Datenmanagement und umfassen Elemente wie Verleiher, Mitarbeiter, Herausgeber, Link (z. B. auf ein Inhaltsverzeichnis), ISBN-Nummer (bei Büchern), Abgabedatum (z. B. bei Dissertationen), Online seit, Änderungsdatum, Verfügbar bis, Verweise, Teil von (übergeordnete Website), Version von, Copyright, Zugangsrechte, Entstehungsprozess, Dateiformat, Quelle usw. Administrative Metadaten können auch Informationen über Rechte, Reproduktion, Selektionskriterien, Logs usw. oder auch Informationen über Rechnungsstellung und Vereinbarungen für die Langzeitaufbewahrung digitaler Objekte beinhalten. Die Grenzen zu anderen Kategorien sind oft fließend.

Technische Metadaten sichern und dokumentieren als Spezialform von administrativen Metadaten die Qualität und die Unversehrtheit eines digitalen Objekts. Anhand dieser Informationen kann die Integrität von Objekten geprüft werden. Zu ihnen zählen Elemente wie Format, Ressource, Grösse, Material, Kompressionsfaktor (bei digitalen Bildern) usw. Weiter können technische Metadaten hardware- und softwarespezifische Informationen beinhalten, aber auch formatspezifische Informationen verwalten wie z. B. Angaben über Formatmigrationen. Im Bereich der digitalen Bilder gehören die Kameradaten zu dieser Kategorie (Exif-Daten). Diese werden bei der Aufnahme automatisch in die Bild-Datei geschrieben.

3.2 FAZIT

Während Metadaten in der analogen Welt oft vom beschriebenen, physischen Objekt getrennt sind,²⁴ können diese im digitalen Kontext sowohl neben dem Objekt verwaltet (z. B. in Inventar-Datenbanken) als auch im Objekt selbst abgelegt werden. Als zusätzliche Sicherheitsmassnahme eignet sich die Anwendung beider Möglichkeiten. Für die Objektbeschreibung ist ein bekannter und weit verbreiteter Metadatenstandard zu wählen, der die benötigten Metadatenkategorien unterstützt und die Nachvollziehbarkeit des digitalen Objekts über die aktive Nutzung hinaus im digitalen Langzeitarchiv sicherstellt.

4 METADATENSTANDARDS

²⁵ Vgl. Woodley, M. S. (2005).

²⁶ TASI, *Metadata Standards and Interoperability* (vgl. Bibliografie). Die TASI stellen weitere «Advice Papers» kostenlos zur Verfügung.

²⁷ Es handelt sich um den ISO-Standard 15836. Weitere Standards befinden sich in der Entwicklung.

²⁸ Eine Übersicht der erwähnten Standards befindet sich im Anhang.

²⁹ Alle von der Library of Congress entwickelt (vgl. Abkürzungen im Abkürzungsverzeichnis).

³⁰ Diese nicht weit verbreiteten Schemata sind keine echten Akronyme: vgl. SPECTRUM (*The UK Museum Documentation Standard*) und MIDAS (*UK Historic Environment Data Standard*).

³¹ Im Bereich der Schriftgutverwaltung wird ein Objekt nicht selten von drei Schemata begleitet (METS, EAD, PREMIS).

³² Eine «Verflachung», auch «dumbing-down» genannt, kann z. B. die Zusammenführung von zwei Elementen wie Vorname und Nachname in ein einzelnes Element Person sein.

Metadatenstandards und -schemata unterstützen neben der Beschreibung digitaler Objekte einen effizienten und sinnvollen Informationsaustausch und dies über unterschiedliche Rechner, Netzwerke und Betriebssysteme hinweg.²⁵ Diese als Interoperabilität bezeichnete Eigenschaft soll die gemeinsame Nutzung unterschiedlicher Datenbestände und Sammlungen erleichtern. Die Begriffe Standard und Schema werden in der Regel nicht unterschieden. Diese werden mehrheitlich synonym gebraucht und beziehen sich im Allgemeinen nur auf die formale Strukturierung der Metadatenelemente.

Ein Standard ist eine von einem internationalen Normierungsgremium genehmigte Norm oder Regelung und hat verbindlichen Charakter (De-iure-Standard). Ein Schema ist in diesem Kontext eine durch seine Verbreitung entstandene Norm (ein De-facto-Standard).

Für die Beschreibung digitaler Objekte existieren viele Standards (oder Schemata), die nicht nur wegen ihrer akronymischen Benennungen die Auswahl erschweren. Eine Schwierigkeit liegt in der Vermittlung einer einfach nachvollziehbaren Übersicht, die trotz guter und prägnanter Beschreibungen den «Laien» zu Beginn abschrecken kann. In diesem Sinne liefert die Kategorisierung der TASI eine übersichtliche und sinnvolle Einführung.²⁶

Wie in den meisten Übersichten wird Dublin Core auch in der TASI-Übersicht als erster Standard und als Beschreibungsschema für Bilder erwähnt. Dublin Core, ursprünglich für die Beschreibung von Websites konzipiert, wird nicht zuletzt seiner Einfachheit wegen am häufigsten eingesetzt und ist einer der wenigen zertifizierten ISO-Standards.²⁷ Als weitere «Bildschemata» werden VRA Core, CDWA und SEPIADES²⁸ erwähnt.

Zu den «gemeinschaftsspezifischen» Standards zählen MARC, MARCXML und MODS²⁹ für Bibliotheken, SPECTRUM, MIDAS³⁰ für Museen fast ausschliesslich im englischen Kontext, ISAD(G), EAD, DACS für Archive und TEI für die Auszeichnung von wissenschaftlichen Texten.

Zu den aufgabenspezifischen Schemata zählen NISO (technische Metadaten für digitale Bilder), PREMIS (Langzeitarchivierung digitaler Daten), METS (Container und Paket für andere Schemata), LOM (Lernressourcen) und OAI-PMH (Interoperabilität mit anderen Standards).

Zu den «Bild-Metadaten» gehören Exif, IPTC und XMP. Diese Informationen werden direkt in die Bild-Datei (TIFF und JPEG) eingetragen und können somit ohne zusätzlichen Aufwand mit dem Bild übermittelt werden.

Quintessenz dieser Kategorisierung und Aufzählung ist die Feststellung, dass ein einziger Standard offenbar nicht genügt, um ein digitales Objekt in seiner Ganzheit zu beschreiben.³¹ In Zukunft könnten – hoffentlich – Vereinfachungen und Verflachungen einzelner Elemente die Beschreibungen erleichtern.³²

³³ «PREMIS defines *preservation metadata* as information a repository uses to support the digital preservation process [...]».

³⁴ «[...] 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.»

Die meisten bekannten Metadatenstandards beschränken sich weiterhin auf die Beschreibung digitaler Objekte, die vor allem das gezielte Auffinden erleichtern soll. Diese rein synchrone Betrachtungsweise, die im alltäglichen Umgang mit digitalen Objekten durchaus Sinn macht, kann in naher Zukunft nicht nur zu technischen Schwierigkeiten, sondern auch zu Problemen bei der Interpretation des Kontextes führen.

Seit einigen Jahren werden die Entwicklungen im Bereich der Metadaten für die Langzeiterhaltung (preservation metadata) mit zunehmendem Interesse verfolgt. Neben Cedars und PADI hat vor allem die PREMIS Arbeitsgruppe das Thema am ausführlichsten behandelt. In ihrem Abschlussbericht, der auch ein ausführliches Data Dictionary vorstellt, definiert PREMIS Metadaten für die Langzeiterhaltung als:

Information, die von einer Ablage [einem digitalen Archiv] verwendet wird, um den digitalen Archivierungsprozess zu unterstützen (PREMIS 2005).³³

Diese funktionale Sicht umfasst die verschiedenen Ausprägungen von Metadaten (beschreibende, strukturelle usw.).

Etwas diffuser wird der Begriff Kern-Metadaten definiert:

Metadaten, die unter allen Umständen benötigt bzw. vorhanden sein müssen, wobei «Kern» nicht unbedingt obligatorisch bedeutet.

Für diesen Widerspruch wird das Argument vorgebracht, dass bei gewissen Ausnahmen «semantische Einheiten» (also Metadatenfelder) optional sein können.³⁴ Diese aus dem Kontext herausgenommene Aussage zeigt erneut auf, dass noch vieles unklar ist und dass weiterhin nach praktikablen Lösungen gesucht wird.

Nicht zuletzt wegen der Komplexität des Themas werden wichtige Entscheidungen in der Hoffnung aufgeschoben, dass zukünftige kommerzielle Lösungen das Problem ganzheitlich angehen und dabei technische und inhaltliche Aspekte auf einen Nenner gebracht werden können.

³⁵ Deshalb empfiehlt es sich, für Konversionen kommerzielle Softwareprogramme einzusetzen (z. B. Photoshop).

4.1 FAZIT BILD-METADATEN

Für digitale Bilder existieren drei Typen von «Metadaten», die in Bilddateien (TIFF, JPEG usw.) verwendet und ausgetauscht werden können (wichtig bei Bildern: Metadaten dürfen bei Formatkonversionen nicht verloren gehen!).³⁵

1. Exif: Technische Informationen über Kamera- oder Scanner-Aufnahmen. Diese Daten werden automatisch erzeugt und durch Benutzer kaum wahrgenommen oder etwa verändert. Aus der Sicht des Nutzers/der Nutzerin ist das wichtigste Feld wohl das Datum der Bilderzeugung (Creation Date).
2. IPTC: Möglichkeit, einfache Beschreibungen in Bild-Dateien zu integrieren. Vorteile sind die weite Verbreitung und die Unterstützung durch viele Bildbearbeitungsprogramme. Es gibt genügend gute Softwarepakete, die auch grosse Bildmengen verarbeiten können. Ein Nachteil ist, dass die Felder für die Pressefotografie konzipiert sind und deshalb oft dazu «missbraucht» werden, um andere Informationen als die vorgesehenen zu speichern, was zu Verwirrung führen kann (vgl. Kap. 5). Ein weiterer Nachteil sind die vergleichsweise kurzen Feldlängen, die keine ausführlichen Bildbeschreibungen zulassen.
3. XMP: Um Nachteile von IPTC zu umgehen hat Adobe XMP mit vielen Erweiterungen definiert. XMP funktioniert nur mit Adobe Software gut, auch wenn immer mehr Softwarepakete diese «Plattform» unterstützen. Es ist aber weiterhin viel Handarbeit nötig und es existiert zurzeit noch keine Datenbank-Einbindung, was jedoch in Zukunft möglich sein sollte.

5 ANWENDUNGSBEISPIEL

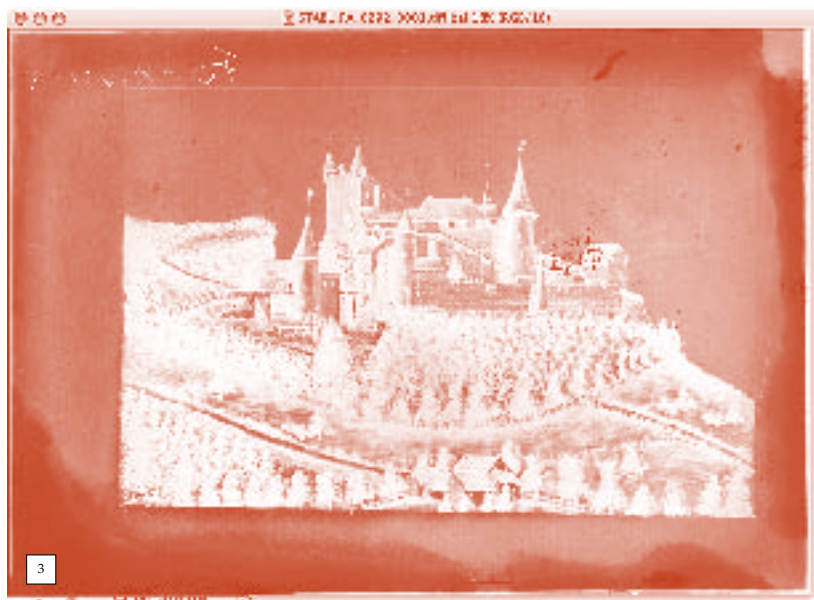
³⁶ *Das Imaging and Media Lab bedankt sich beim Staatsarchiv Baselland für die Erlaubnis, einen Teil seines Bildbestands für Illustrationszwecke nutzen zu dürfen.*

³⁷ *Die Information befand sich in der mitgelieferten Excel-Datei unter der Rubrik «Source».*

Im folgenden Beispiel wird aufgezeigt, wie digitale Bilder auf einfache Weise mit Metadaten versehen werden können. Die hier verwendeten Bilder wurden dem IML vom Staatsarchiv Baselland zwecks Digitalisierung zugestellt. Zusammen mit den analogen Originalen lieferte das Archiv auch die Bildbeschreibungen (in einer Excel-Datei).³⁶

Schritt 1

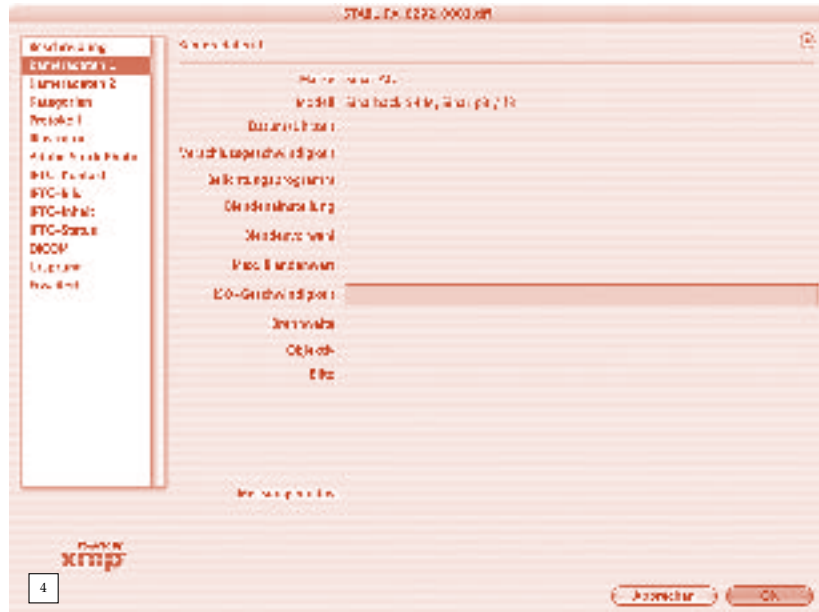
Die Bilder wurden mit einer SINAR-Digitalkamera digitalisiert und bereits mit dem korrekten Dateinamen nach vorgegebener Namenskonvention benannt.³⁷



Die technischen Metadaten (Exif-Daten) wurden während der Digitalisierung automatisch in die Datei eingetragen. Diese Informationen können z. B. in Photoshop angezeigt werden. In diesem Beispiel werden nur Marke und Kameramodell angezeigt, bei Aufnahmen ab Digitalkamera hat man normalerweise mehr Informationen.

³ Beispiel: STABL_PA_6292_0001.TIFF

³⁸ Mit CaptionWriter der Firma Comnet Software können IPTC-Beschreibungen für eine ganze Bildsammlung importiert werden. Adobe Bridge erlaubt nach Auswahl der entsprechenden Bilder zurzeit nur eine halbautomatische Beschreibung der IPTC-Felder oder nur das Importieren von Metadaten für ein einziges Bild.



Schritt 2

Nach Abschluss der Digitalisierung wurden hochauflösende JPEG-Versionen ab TIFF-Vorlage erstellt. Diese können als Arbeitskopien oder für die Distribution genutzt werden. Dabei bleibt die TIFF-Masterdatei (das «Original») unangetastet und kann für die Produktion weiterer Kopien verwendet werden.

Schritt 3

Nach der Produktion der JPEG-Dateien erfolgte die Definition der beschreibenden Metadaten, die der Auftraggeber in einer Excel-Datei mitgeliefert hatte. Möglicherweise stammen diese aus einer Datenbank. Die Feldnamen entsprachen bereits den IPTC-Feldnamen. Dank der eindeutigen Zuweisung des Dateinamens (z. B. STABL_PA_6292_0001.TIFF im Feld ObjectName, vgl. Abbildung oben) war es möglich, eine automatisierte Metadaten-Eingabe durchzuführen. Bei mehr als 1000 Bildern konnte damit sehr viel Zeit eingespart werden. Zudem konnte dadurch auch die Fehlerquote, die bei manueller Eingabe unweigerlich entsteht, nicht nur reduziert, sondern ganz vermieden werden. Für die automatisierte Eingabe wurde das Programm CaptionWriter eingesetzt.³⁸

³⁹ Diese Beschränkung soll mit IPT-C4XMP in Zukunft aufgehoben werden, setzt aber voraus, dass die benutzte Softwarelösung das XMP metadata framework von Adobe unterstützt (vgl. <http://www.iptc.org/IPTC4XMP/>).

№	Objektname	Georeferenz	City	Province	Country	SpecialInstructions	Caption	Media
1	STABL_PA_6292_0001	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0001_0001	Archivierung
2	STABL_PA_6292_0002	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0002_0001	Archivierung
3	STABL_PA_6292_0003	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0003_0001	Archivierung
4	STABL_PA_6292_0004	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0004_0001	Archivierung
5	STABL_PA_6292_0005	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0005_0001	Archivierung
6	STABL_PA_6292_0006	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0006_0001	Archivierung
7	STABL_PA_6292_0007	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0007_0001	Archivierung
8	STABL_PA_6292_0008	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0008_0001	Archivierung
9	STABL_PA_6292_0009	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0009_0001	Archivierung
10	STABL_PA_6292_0010	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0010_0001	Archivierung
11	STABL_PA_6292_0011	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0011_0001	Archivierung
12	STABL_PA_6292_0012	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0012_0001	Archivierung
13	STABL_PA_6292_0013	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0013_0001	Archivierung
14	STABL_PA_6292_0014	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0014_0001	Archivierung
15	STABL_PA_6292_0015	Landesamt	Ortenaukreis	DL	Schweiz	Spezialanweisungen	STABL_0015_0001	Archivierung

Die Definition erfolgte getrennt für die TIFF- und für die JPEG-Version. Dieses Vorgehen hat einen Grund. Nicht selten gehen bei Formatkonversionen Informationen aus IPTC-Feldern verloren (z. B. von TIFF nach JPEG). Auch ist bei IPTC die Zahl der beschreibbaren Felder vorgegeben, die Feldlänge beschränkt und je nach Anwendung möglicherweise knapp bemessen.³⁹ Diese Informationen können ebenfalls in Photoshop oder in vielen anderen Bildbearbeitungsprogrammen visualisiert werden.

The screenshot shows the 'IPTC-Subheader' section of the metadata editor. It contains several text input fields:

- Titel:** Landesamt Ortenaukreis
- Bezeichnung:** Ortenaukreis
- Ortsbezeichnung:** Ortenaukreis
- IPTC-Subheader Code:** (empty)
- Wörter für Beschreibung:** (empty)

At the bottom left, there is a small red box with the number '6'. At the bottom right, there are 'Abbrechen' and 'OK' buttons.

⁴⁰ Beispiele für Softwareprodukte, die Bilder verwalten, finden sich auf der TASI-Webseite «Systems for Managing Image Collections»: <http://www.tasi.ac.uk/adoice/delivering/ims2.html>

⁴¹ Das Beispiel soll auch aufzeigen, dass die Definition von Mappings («Konkordanzen») zwischen Schemata, die für unterschiedliche Anwendungen konzipiert worden sind, schwierig ist und dass oft keine Entsprechungen dafür vorhanden sind.

⁴² Praktische Beispiele für Feldinhalte, die nicht den ursprünglichen Feldbeschreibungen entsprechen.

⁴³ Zu klären wäre, ob diese Information als <subject> im Header oder im Beschreibungsteil definiert werden soll.

Eine rasche Übersicht über den Bildbestand erhält man z. B. mit Adobe Bridge.⁴⁰



Schritt 4

Nach der Produktion der JPEG-Arbeitskopien und der Erfassung der Metadaten wurden die Bilder dem Auftraggeber zusammen mit den analogen Originalvorlagen übermittelt und der digitalisierte Bildbestand für die Langzeitarchivierung vorbereitet, die in Kapitel 7 behandelt wird.

In diesem Anwendungsbeispiel wurde nur aufgezeigt, wie Metadaten in Bilddateien geschrieben werden. Es ist auch möglich, Meta-informationen in einer externen Datei zu verwalten. In der folgenden Tabelle wird ansatzweise demonstriert, wie dieselben Informationen z. B. in Dublin Core und EAD abgelegt werden könnten.⁴¹

IPTC	Dublin Core	EAD	Information Beispiel
Titel (Title)	dc.title	<unittitle>	STABL_PA_6292_0001 ⁴²
Überschrift (Headline)	dc.subject	<subject> ⁴³	Schloss Farnsburg ...
Anbieter (Credit)	dc.creator	<author>	Seiler-Schaub
Erstellungsdatum (Date created)	dc.description? dc.date?	<unitdate?>	1895?
...

6 KERNFELDER

⁴⁴ Vgl. <http://www.iptc.org/IPTC4XMP/>

⁴⁵ Bei dieser beschränkten Auswahl von Attributen zeigt sich wiederum das Problem der Interoperabilität. Die Semantik der einzelnen Felder kann divergieren, was bei einer künftigen Übernahme in ein neues Schema Interpretationsprobleme verursachen kann.

⁴⁶ In diesem Fall ist nicht klar, welches Datum gemeint ist (aktuell oder historisch?).

Im Bereich der digitalen Bilder bleibt die Definition von IPTC-Metadaten eine empfehlenswerte Vorgehensweise. Auch wenn Metadaten in einer externen Anwendung verwaltet werden können (z. B. in einer Datenbank), empfiehlt es sich, die gleichen Informationen auch in die entsprechenden Bild-Dateien abzulegen. Sollte in Zukunft der Verweis zwischen Anwendung und Bild nicht mehr bestehen, können anhand der im Bild integrierten Beschreibungen Inhalt und Kontext rekonstruiert werden. Seit 2004 arbeiten IPTC und Adobe an einem gemeinsamen IPTC-Kernschema für XMP (IPTC4XMP), das Informationen zwischen Standard-IPTC Feldern (Ver. 1.0) und dem neuen XML-basierten Schema in beide Richtungen synchronisiert.⁴⁴ Die Definition von IPTC-Metadaten bleibt damit ein sinnvolles und für die Zukunft gültiges Vorgehen.

Für die Definition von Kernfeldern können die Empfehlungen der RLG Descriptive Metadata Guidelines (2005) als Grundlage dienen. Im Bereich der digitalen Bilder können folgende Felder den Kontext weitgehend beschreiben:

	RLG (Vorschlag)	IPTC	Dublin Core
Identifikation	Unique ID	Object Name?	dc.identifier ⁴⁵
Beschreibung	Work Description	Caption	dc.description
Sammlung	Collection Title	Headline	dc.title
Thema	Collection Subjects	Category	dc.subject
Stichwort	?	Keywords	dc.subject
Urheber / Autor	«Creator»	Credit?	dc.creator
Datum	Date Range?	Create Date	dc.date? dc.coverage? ⁴⁶

Kernfelder sollten grundsätzlich auf ein aussagekräftiges Minimum beschränkt werden. Bei «gründlichen» Objektbeschreibungen mit vielen Attributen besteht das Risiko, dass aus Zeitmangel ohnehin nur die «wichtigsten» Felder beschrieben werden.

7 DIGITALE ARCHIVIERUNG

Nach Abschluss der digitalen Erschliessungsarbeiten (Digitalisierung, Definition von Metadaten, Produktion von Arbeitskopien) stellt sich die Frage nach dem Langzeiterhalt der neu produzierten digitalen Daten. Das grundsätzliche Problem liegt in der Speicherung digitaler Informationen, die in einem vereinfachten Schichtenmodell wie folgt aussieht (Gschwind 2007):

Information, Metadaten	Dossiers, Dokumente, Fotos, E-Mail usw.
Software, Dateiformate	Anwendungsprogramme, Dateien
Hardware, Datenträger	Prozessor, Speicher, Magnetband, optische und magnetische Festplatten (hard disks)

Die unterste Schicht (Hardware Schicht) ist aus verschiedenen Gründen äusserst instabil und kurzlebig: technologischer Wandel, obsoleter Datenträger, Datenträgerverschleiss usw. sind die Hauptprobleme.

Diese Instabilität zeigt sich auch auf der mittleren Schicht (Software Schicht), z. B. bei neuen Formaten. Der Wechsel erfolgt langsamer, ist aber gleichermassen problematisch.

Bei der obersten Schicht (Informationsschicht) ergibt sich grundsätzlich kein rein digitales Problem. Fehlt jedoch die unterste Schicht, haben Dateien und Metadaten keine Relevanz mehr. Die Information ist unwiderruflich verloren.

Das grundsätzliche Problem der dauerhaften Archivierung digitaler Daten liegt in der Informationsspeicherung. Information wird mit Hilfe von Computern in Binärcode umgewandelt und auf digitalen Datenträgern gespeichert (Aufbewahrung des Bitstroms). Binärcode kann bei richtiger Handhabung beliebig oft und ohne Informationsverlust kopiert werden. Aus den oben genannten Gründen muss dieser Kopiervorgang in regelmässigen Zeitintervallen, in der Regel alle fünf Jahre oder früher, erfolgen. Dieser Vorgang hat einen kompletten Austausch der Technologie zur Folge, wobei die Informationen auf neue Datenträger kopiert werden. Dies muss ausnahmslos und ohne Unterbruch geschehen. Eine Migration verursacht hohe Fixkosten und benötigt das entsprechende Fachwissen, damit Daten nicht verloren gehen.

Der nachfolgend beschriebene Archivierungsprozess ist das Ergebnis aus mehrjähriger Forschungstätigkeit und praktischer Anwendung des Imaging and Media Lab im Bereich der digitalen Archivierung von Bildern. Das Vorgehen kann sowohl bei der Erst-Archivierung als auch bei nachfolgenden Archivierungen und Migrationen angewendet werden. Es handelt sich um eine generische Vorgehensweise, die an spezifische Bedürfnisse angepasst werden kann.

⁴⁷ Eine ausführlichere Beschreibung des Archivierungsprozesses kann auf der Website des IML eingesehen werden: <http://www.iml.unibas.ch/index?content=28&news=1>

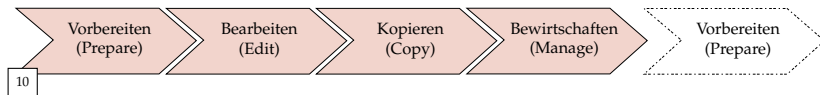
⁴⁸ Entspricht im Grundsatz einem Ablieferungspaket im OAIS-Modell: Submission Information Package (SIP), vgl. OAIS Reference Model in der Bibliografie.

⁴⁹ Vgl. Message-Digest Algorithm 5 (MD5): http://de.wikipedia.org/wiki/Message-Digest_Algorithm_5

⁵⁰ Zur Veranschaulichung werden für dieses Beispiel absichtlich JPEG-Dateien benutzt, weil die Prüfsummenberechnung und Prüfung für grössere TIFF-Bildbestände viel Zeit beansprucht (vgl. Kap. 7.1.4, Unterabschnitt Bewirtschaften).

7.1 ARCHIVIERUNGSPROZESS

Der Archivierungsprozess umfasst die vier Teilprozesse Vorbereiten, Bearbeiten, Kopieren und Bewirtschaften:⁴⁷



7.1.1 VORBEREITEN

Die Vorbereitung beginnt mit der Speicherung der neu erstellten oder aus analogen Quellen produzierten digitalen Daten. Dieser Teilprozess wurde im Anwendungsbeispiel bereits veranschaulicht (Kapitel 5, Schritt 1). Es empfiehlt sich, für alle Datensammlungen eine für die Datennutzer nachvollziehbare Ablagestruktur zu definieren (Ordner- bzw. Verzeichnishierarchie). Zur Vorbereitung gehören unter anderem folgende Schritte:

- Bewertung und Selektion
- Archivpaket definieren (Verzeichnis, Anzahl Dateien)⁴⁸
- Datenmenge ermitteln (beeinflusst die Kosten, vgl. Selektion)
- Verzeichnisstruktur dokumentieren
- Wahl des Archivmediums
- Archivier-Infrastruktur prüfen
- Archivier-Zeitpunkt festsetzen (eingeschränkte Nutzung IT-Infrastruktur)

7.1.2 BEARBEITEN

Bei diesem Teilprozess werden einerseits alle für die Archivierung benötigten Metadaten definiert (beschreibende und administrative Metadaten; vgl. Kap. 5, Schritt 3) und andererseits für alle Dateien Prüfsummen berechnet und in einer Textdatei abgelegt, die zusammen mit den Dateien archiviert wird.

Die wichtigsten Schritte sind:

- Dateien zusammenfassen (Thema, Verzeichnisinhalt, Anzahl)
- IPTC Metadaten definieren und bearbeiten (einzeln, gruppenweise, automatisch)
- Beschreibungen prüfen (Qualitätskontrolle, Vollständigkeit, Klassifikationsschema?)
- Prüfsummenalgorithmus wählen (z. B. MD5⁴⁹)
- Prüfsummen für Dateiauswahl bzw. Archivier-Paket berechnen
- Prüfsummendatei im gleichen Verzeichnis ablegen (Vergleichsbasis für zukünftige Integritätsprüfungen, wird zusammen mit den Bilddaten archiviert)

Prüfsummen erlauben eine einfache Datenintegritätsprüfung. Anhand von Prüfsummen kann mit einem Vergleich festgestellt werden, ob eine Datenübertragung oder -speicherung korrekt erfolgt ist (z. B. ohne Übertragungsfehler)⁵⁰.



11



12

¹⁰ Archivierungsprozess (IML, Universität Basel)

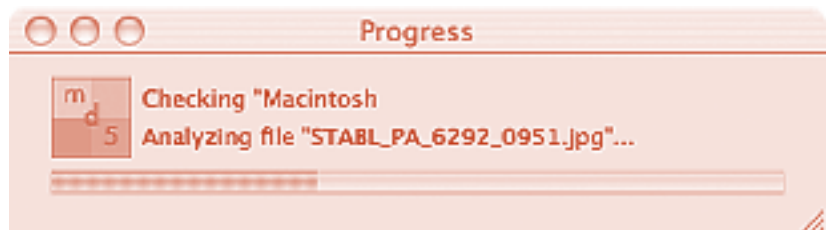
¹¹ Prüfsummenberechnung (z. B. mit CheckSum+)

¹² Inhalt der Prüfsummendatei

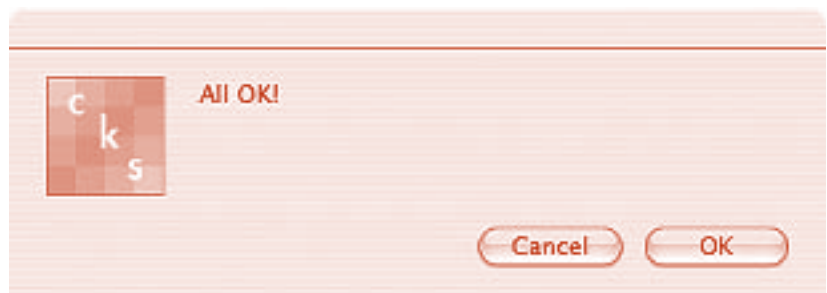
7.1.3 KOPIEREN

Mit diesem Teilprozess werden die zu archivierenden Dateien in das digitale Archiv transferiert, anschliessend werden die Prüfsummen neu berechnet und mit jenen in der mit archivierten Textdatei verglichen (vgl. Teilprozess in Kap. 7.1.2 Bearbeiten). Die wichtigsten Schritte sind:

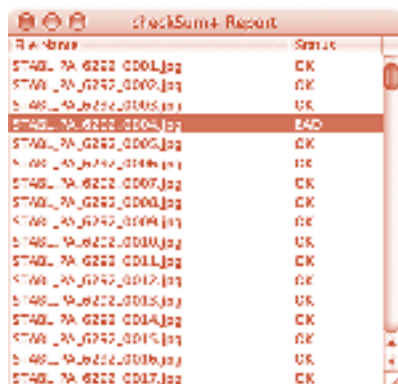
- Dateien in das digitale Archiv kopieren
- Mindestens drei Kopien anlegen (später an verschiedenen Orten aufbewahren)
- Nach jeder Kopie ins Archiv Prüfsummen neu berechnen und vergleichen



13



14



File Name	Status
STABL_PA_6292_0001.jpg	OK
STABL_PA_6292_0002.jpg	OK
STABL_PA_6292_0003.jpg	OK
STABL_PA_6292_0004.jpg	OK
STABL_PA_6292_0005.jpg	OK
STABL_PA_6292_0006.jpg	OK
STABL_PA_6292_0007.jpg	OK
STABL_PA_6292_0008.jpg	OK
STABL_PA_6292_0009.jpg	OK
STABL_PA_6292_0010.jpg	OK
STABL_PA_6292_0011.jpg	OK
STABL_PA_6292_0012.jpg	OK
STABL_PA_6292_0013.jpg	OK
STABL_PA_6292_0014.jpg	OK
STABL_PA_6292_0015.jpg	OK
STABL_PA_6292_0016.jpg	OK
STABL_PA_6292_0017.jpg	OK

15

- Bei Fehlern Ursache ermitteln (Übertragungsfehler, Medium usw.)
- Fehler unverzüglich korrigieren!

Sind die Prüfsummen der ursprünglichen Dateien mit denen im Archiv identisch, entspricht die Archivkopie der «Originaldatei» (erfolgreiche Integritätsprüfung). Sind die Summen nicht identisch, liegt ein Fehler vor. Dieser kann mehrere Ursachen haben: unvollständige Übertragung, beschädigtes Archivmedium, Fehlmanipulation usw. In diesem Fall muss die fehlerhafte Datei ersetzt werden. Die Wichtigkeit redundanter Datenhaltung kann in diesem Zusammenhang nicht oft genug betont werden.

⁵¹ Der hier dokumentierte Archivierungsprozess deckt nur die ersten beiden Komponenten ab (Ablieferung und Archivierung). Die Nutzung wurde bewusst ausgenommen, da diese prinzipiell nicht zur Archivierung gehört.

⁵² Für die Verwaltung digitaler Archive werden Softwarelösungen entwickelt. Im OpenSource-Bereich erhalten Fedora (<http://fedoraproject.org/>) und Dspace (<http://www.dspace.org/>) zunehmend Aufmerksamkeit.

7.1.4 BEWIRTSCHAFTEN

Nach der Übertragung der Dateien ins digitale Archiv ist der technische Teil der Archivierung abgeschlossen. Die Bewirtschaftung des Archivs umfasst verschiedene Führungsaufgaben, die jederzeit nachvollziehbar und nicht personenabhängig sein sollten. Zur Bewirtschaftung eines digitalen Archivs gehören unter anderem folgende Tätigkeiten:

- Archivierte Daten dokumentieren (welche Daten, Ablageort, ...)
- Ordnungsdaten (Art von Daten, Zugriff)
- Periodische Prüflistenprozeduren planen und durchführen (z. B. alle sechs Monate, ist das Medium lesbar?, Prüfsummen neu berechnen und vergleichen)
- Erste / nächste Migration planen
- Migrationsvorhaben rechtzeitig budgetieren
- Aufbewahrungsort definieren (In-House oder extern aufbewahren: Kriterien Preis / Sicherheit / Zugriffszeit)
- Aus- und Weiterbildung des Personals planen und budgetieren

Je nach Datenvolumen kann eine Sicherung Stunden, aber auch Tage dauern. Umso wichtiger ist deshalb die Planung der Sicherungsprozedur. Weiter ist zu bedenken, dass bei korrekt ausgeführten Archivierungsprozessen mehrere Datenträger zwecks Redundanz beschrieben werden müssen. Der Zeitaufwand für die Sicherung ist mit der Anzahl der zu erstellenden Kopien zu multiplizieren. Nach erfolgter Sicherung müssen die Datenträger und die Begleitdokumentation (vgl. Vorbereitung) an verschiedenen Orten abgelegt werden, am besten in verschiedenen Gebäuden, die geografisch weit genug auseinander liegen. Seit einiger Zeit wird empfohlen, digitale Archive nach dem OAIS-Referenzmodell zu verwalten. Das Modell sieht drei Hauptkomponenten vor: die Ablieferung als SIP (Submission Information Package), die Archivierung als AIP (Archive Information Package) und die Nutzung als DIP (Dissemination Information Package).⁵¹ Das OAIS-Referenzmodell legt nur die Rahmenbedingungen fest. Die konkrete technische Umsetzung der Ablieferungs-, Archivierungs- und Nutzungsfunktionalität bleibt den archivierenden Institutionen überlassen.⁵² Weiterführende Informationen können der Originaldokumentation entnommen werden (vgl. Bibliografie).

7.2 FAZIT

Während die Erst-Archivierung einen wertschöpfenden Prozess darstellt, sind die nachfolgenden Migrationen vor allem werterhaltende Prozesse, die wiederkehrende Kosten verursachen. Die finanziellen Mittel für diesen Werterhalt, aber auch für die Infrastruktur, müssen zum Migrationszeitpunkt zur Verfügung stehen. Wird eine Migration aus Mangel an finanziellen Ressourcen ausgelassen, können digitale Daten schnell verschwinden. Dieses Bewusstsein und die daraus folgende finanzielle Verpflichtung müssen vorhanden sein.

8 SCHLUSSFOLGERUNGEN UND AUSBLICK

⁵³ Seit 2003 ist die DCMI ein ISO-Standard (15836-2003) und seit Mai 2007 ein NISO-Standard (Z39.85-2007).

⁵⁴ XMP von Adobe legt Metainformationen standardmäßig als Dublin Core-Elemente ab.

Während sich im Bereich der Digitalisierung erprobte Vorgehensweisen schrittweise konsolidieren, bleibt das Thema Metadaten komplex und oft auch unbefriedigend gelöst. Obwohl viele Schemata existieren (vgl. Kap. 10.3), steht eine Lösung, die digitales Kulturgut einheitlich beschreibt und verwaltet, weiterhin aus. Dennoch ist die Definition von Metadaten unerlässlich, damit digitale Objekte auch in Zukunft auffindbar und begreifbar bleiben. Dies trifft besonders auf digitale Bilder zu, deren Inhalt nicht anhand von textuellen Informationen nacherschlossen werden kann.

Die Definition von Metadaten im digitalen Bild selbst bietet einige Vorteile. Das Objekt beinhaltet alle darin abgelegten Informationen und ist im Prinzip nicht auf die Unterstützung einer Verwaltungssoftware angewiesen, was nicht bedeutet, dass die externe Verwaltung von Information und Kontext nur Nachteile bringt. Für das Tagesgeschäft bietet eine Verwaltungssoftware viele Vorteile, z. B. eine bessere Übersichtlichkeit oder einen schnelleren Zugriff auf ein digitales Objekt. Mit einem solchen Vorgehen vereinfacht sich auch der Archivierungsprozess, indem ein Einzelobjekt beim Kopiervorgang alle enthaltenen Informationen automatisch mit transportiert. Das Ablegen von Informationen in Bild-Dateien hat aber auch Nachteile. Exif- und IPTC-Felder lassen sich einfach verändern oder löschen. Deshalb müssen zusätzliche Vorkehrungen getroffen werden, um digitale Bilder, aber auch allgemein digitale Objekte, vor unerwünschten Manipulationen zu schützen.

Im Bereich der Metadaten-Standards dient Dublin Core weiterhin als Referenzschema. Dublin Core existiert seit 1995 und besteht aus einer überschaubaren Liste von 15 Elementen, die sich z. B. auf einfache Weise auch in HTML-Dokumente integrieren lassen. Die Dublin Core-Metadata-Initiative engagiert sich für interoperable Metadaten und ist nunmehr auch ein De-iure-Standard,⁵³ auf den sich andere Standards beziehen.⁵⁴ Bei der Wahl eines Schemas sollte deshalb darauf geachtet werden, dass die Interoperabilität mit Dublin Core gewährleistet ist. Dublin Core kann als Lingua Franca unter den Metadaten-Standards betrachtet werden. Auch andere Standards sollen mit Lite-Versionen überschaubarer werden (z. B. CDWA Lite, VRA Core, vgl. Kap. 10). Es ist zu erwarten, dass weitere Vereinfachungen folgen werden.

In diesem Sinne wird sich auch die Beschränkung auf eine übersichtliche Menge von Kernfeldern bewähren.

Im Bereich der digitalen Archivierung gehört die Sensibilisierungsphase der Vergangenheit an. Digital archivieren bedeutet, einen bestimmten Zustand auf einem Datenträger ablegen, der bis zur nächsten Migration unverändert an einem sicheren Ort aufbewahrt wird. Das Vorgehen ist klar vorgezeichnet. Damit ergibt sich auch die Verpflichtung einer langfristigen Finanzierung, mit der die Verfügbarkeit von digitalem Archivgut garantiert wird. Die digitale Archivierung ist ein aktiver Prozess, der keine Unterbrüche duldet und konsequente Aktivitäten voraussetzt.

Die Umsetzung der in den vorliegenden Guidelines dokumentierten Vorgehensweise ist ein erster Schritt. Damit digitale Bilder auch in Zukunft verfügbar und begreifbar bleiben, können aus heutiger Sicht einige Empfehlungen abgegeben werden, die in der Checkliste in Kap. 10.1 aufgeführt sind. Die vorliegenden Guidelines sind auch eine Momentaufnahme: Technologien, Software und Standards sind stetigen Veränderungen unterworfen. Deshalb wird es in Zukunft nötig sein, die hier beschriebene Vorgehensweise neuen Rahmenbedingungen und Gegebenheiten anzupassen.

9 BIBLIOGRAFIE

*Stand:
Februar 2008*

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

Deutsche Nationalbibliothek, Schlagwortnormdatei:
<http://www.ddb.de/standardisierung/normdateien/swd.htm>

Gschwind R. 2007: Braucht es in der Schweiz ein Kompetenzzentrum für die Langzeitarchivierung digitaler Daten? Bern: 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, 2007: 411–417.

Menne-Haritz A. 2005: METS: Überblick und Anleitung (übers. METS: An Overview und Tutorial):
http://www.loc.gov/standards/mets/METSOverview.v2_de.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://wikicatalog.pbwiki.com/f/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 Photographpy, 4th Ed., Elsevier Focal Press, 2007: 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>

Verein Schweizerischer Archivarinnen und Archivare (VSA), Katalog wichtiger, in der Schweiz angewandter archivischer Normen:
http://www.vsa-aas.org/uploads/media/Normenkatalog_Version1-1_20070901.pdf

10 ANHANG

Diese Checkliste fasst die wichtigsten Punkte zusammen und ist als Orientierungshilfe konzipiert.

10.1 CHECKLISTE

Allgemein	
Metadaten	
Relevante IPTC-Felder in digitalen Bildern mit Informationen versehen	
Spezielle Bedeutung von Attributen (Abweichung der ursprünglichen Benennung) dokumentieren (insbesondere bei IPTC-Feldern)	
Kernfelder auf ein überschaubares Minimum beschränken	
...	
Metadatenschema	
Dublin Core als Referenz verwenden oder als Orientierungsschema in Betracht ziehen (Interoperabilität, «Felder-Konkordanz» für Schemawechsel)	
Weiterentwicklung von Metadatenstandards und -schemata verfolgen	
...	
Technologische Entwicklung verfolgen (technology watch)	
Entwicklung von Speichermedien verfolgen	
Entwicklung von Dateiformaten verfolgen	
Formatmigrationen prüfen (Qualität, Informationsverlust, Reversibilität usw.)	
...	
Digitale Archivierung (Prozess)	
Vorbereitung	
Bewertung und Auswahl	
Archivpaket(e) definieren	
Datenmenge ermitteln	
Verzeichnisstruktur dokumentieren	
Wahl des Archivmediums	
Archivier-Infrastruktur prüfen	
Archivier-Zeitpunkt festsetzen	
...	
Bearbeiten	
Dateien zusammenfassen (thematisch, nach Verzeichnisinhalt, Anzahl)	
IPTC-Metadaten definieren und bearbeiten	
IPTC-Beschreibungen prüfen	
Prüfsummen für Archivier-Paket(e) berechnen	
Prüfsummendatei mit den Archivdaten ablegen (gleiches Verzeichnis)	
...	

Kopieren	
Dateien auf die Archivressource kopieren (digitales Archiv)	
Mindestens drei Kopien anlegen (Redundanz)	
Nach jedem Kopiervorgang Prüfsummen neu berechnen und vergleichen	
...	
Bewirtschaften	
Finanzierung langfristig gewährleisten	
Migrationsvorhaben rechtzeitig budgetieren	
Kostenentwicklung verfolgen (Zunahme der Datenmenge)	
Archivier-Infrastruktur warten und an den technologischen Wandel anpassen	
Archivierte Daten dokumentieren (welche Daten?, Ablageort usw.)	
Ordnungsdaten (Art von Daten, Zugriff)	
Geplante periodische Prüfprozeduren durchführen (Medium, Prüfsummen)	
(Erste) Nächste Migration planen	
Daten in geografisch getrennten Gebäuden aufbewahren	
Aufbewahrungsort definieren (In-House oder extern aufbewahren: Kriterien Preis / Sicherheit / Zugriffszeit)	
Aus- und Weiterbildung des Personals budgetieren, planen und durchführen	
...	

10.2 RLG KERNFELDER (SUMMARY OF CORE FIELDS)

Eintrag Sammlung (Collection Record)

Grunddaten (Base-Line)

Collection Title (Titel [der Sammlung])

Collection Contributor (Beteiligte/r an der Sammlung)

Collection Date Range (Datumsauswahl der Sammlung)

Collection Narrative Description (Beschreibung der Sammlung)

Collection Work Type (Werk[-Form] der Sammlung)

Collection Subjects (Topics, People, Places):
(Bereiche der Sammlung [Themen, Personen, Orte])

Eintrag Objekt (Individual Record)

Grunddaten (Base-Line)

Creator (Urheber)

Work Type (Werk[-Form])

Work Title (Titel des Werks)

Creation Date (Erstellungsdatum)

Unique ID (Einzigartige ID)

Pointer/s to Surrogate/s (Hinweis/e auf Surrogat/e)

«Mehrwert-Information» (Value-added)

Place of Creation (Erstellungsort)

Subjects (Topics, People, Places) (Bereiche [Themen, Personen, Orte])

Work Description (Werkbeschreibung)

Bonus

Sequence / Hierarchy of Surrogates (Multi-part work) (Folge /
Hierarchie der Surrogate)

Caption for Surrogates (Multi-part work) (Beschriftung Surrogate)

Measurement (Masse)

Materials and Techniques (Material und technische Methoden)

Detailinformationen: RLG (2005: 13)

10.3 ÜBERSICHT: METADATENSTANDARDS UND -SCHEMATA

Institutionen, die sich schon sehr früh mit dem Thema Metadaten auseinandersetzten – insbesondere Bibliotheken –, taten dies aus dem Bedürfnis heraus, ihre neu entstandenen digitalen Findmittel (später auch ihre digitalen Bestände) zu beschlagworten. Bald erkannte man, dass «digitale Kataloge» nicht nur einen effizienteren Zugriff auf Informationen ermöglichten, sondern auch beim Datenaustausch gegenüber konventionellen Wegen Vorteile boten, was vor allem für die Forschung interessant wurde. Dieser Austausch war schneller und bei richtiger Vorgehensweise auch preiswerter.

Eine richtige Vorgehensweise bedeutet, dass ein Austauschformat vorliegt, das die Sender und Empfänger der Daten codieren bzw. decodieren können. Dies ist die Grundbedingung für einen erfolgreichen Datenaustausch und auch der Ursprung von Austauschformaten (wie z. B. RTF), die Textformatierungen beibehalten. Im Kontext von Metadaten wird vermehrt der Begriff Interoperabilität benutzt. Diese ermöglicht nicht nur den Datenaustausch, sondern unterstützt auch die Interaktion zwischen unterschiedlich konzipierten Systemen.

In der folgenden, nicht abschliessenden Aufzählung werden weit verbreitete Metadatenstandards, -schemata und -initiativen kurz vorgestellt.

Dokumentation:
<http://dublincore.org>

10.3.1.1 DUBLIN CORE

Dublin Core ist das am weitesten verbreitete Metadatenschema und besteht aus 15 Elementen: Titel (title), Urheber (creator), Thema und Stichwörter (subject), Beschreibung (description), Verleger oder Herausgeber (publisher), Beteiligte (contributors), Datum (date), Objektart (type), Format (format), Identifikation (identifier), Quelle (source), Sprache (language), Beziehungen (relation), Gültigkeitsbereich (coverage) und Rechtliches (rights).

Dublin Core-Datenelemente können als HTML-Metatags, XML-, RDF-Tags oder als Feldattribute in Datenbanken definiert werden. Ursprünglich für die Beschreibung von digitalen Dokumenten vorgesehen, wird das Schema auch für die Katalogisierung von Museumsobjekten (physische Objekte) eingesetzt.

Dokumentation:
www.vraweb.org

10.3.1.2 VRA (VISUAL RESOURCES ASSOCIATION) / VRA CORE

VRA ist ein Standard für die Beschreibung von Bildern (visuelle Werke) im Kulturbereich. Die Metadatenelemente charakterisieren das Werk und seine Abbildungen hierarchisch und nach Kategorien. VRA Core, das sich auf CDWA und Dublin Core stützt, besteht aus 17 Kernelementen.

Dokumentation:
http://www.getty.edu/research/conducting_research/standards/cdwa/

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

Die CDWA definieren eine Menge von Metadatenelementen für die Beschreibung von Kunstwerken und sind vor allem für den Zugriff und die Recherche konzipiert. Die Objektbeschreibungen sind hierarchisch aufgebaut. Als Alternative zu den ursprünglich 500 und mehr Kategorien und Unterkategorien wurde eine CDWA-Lite-Version entwickelt, deren Kernelemente auch als XML-Schema publiziert worden sind.

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

Werkzeug / Sepiades Tool:
<http://www.knaw.nl/ECPA/sepia/workinggroups/wp5/download.html>

10.3.1.4 SEPIADES (SEPIA DATA ELEMENT SET)

SEPIADES ist 2003 aus dem SEPIA-EU-Projekt hervorgegangen (Safeguarding European Photographic Images for Access). Im Fokus stehen Fotoarchive. Der Ansatz ist hierarchisch und beruht auf dem Verzeichnungsstandard ISAD(G). Zuerst wird für eine Fotosammlung ein Metadateneintrag erstellt, der dann weiter auf Objektgruppen und bei Bedarf auf Einzelobjekte heruntergebrochen wird.

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

10.3.1.5 MACHINE-READABLE CATALOGING (MARC & MARCXML)

Die MARC21-Formate definieren Standards für die Darstellung und die Übermittlung bibliografischer Informationen in maschinenlesbarer Form. Der Datenaustausch erfolgt entweder über die beiden Austauschformate ISO 2709 und ANSI / NISO Z39.2 oder über XML. MARC ist ein «bibliografisches Format» für Bücher, Tonaufnahmen, Bewegtbilder sowie Archivalsammlungen und ist in Bibliotheken und Archiven der am weitesten verbreitete Standard.

Dokumentation:

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

10.3.1.6 UNIMARC & XML UNIMARC

UNIMARC und XML UNIMARC (universal MARC) sind Standards für die Darstellung und Übermittlung von bibliografischen Informationen in maschinenlesbarer Form. UNIMARC spezifiziert eine gemeinsame Basis für den Datenaustausch bibliografischer Informationen für internationale MARC-Formate.

Dokumentation:

www.loc.gov/standards/mods/

10.3.1.7 METADATA OBJECT DESCRIPTION SCHEMA (MODS)

Das Metadata Object Description Schema (MODS) ist ein in XML codiertes Metadatenschema für bibliografische Informationen. MODS beinhaltet eine Untermenge von MARC21-Feldern und positioniert sich zwischen dem einfacheren Dublin Core und der ausführlicheren Beschreibung von MARC21. MODS kommt möglicherweise in Bibliotheken und Archiven zur Anwendung, die bereits MARC21 oder MARCXML nutzen. MODS soll als vereinfachte Version von MARCXML für Menschen besser lesbar sein.

Dokumentation:

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

10.3.1.8 SPECTRUM

SPECTRUM, The UK Museum Documentation Standard, ist ein national und international anerkannter Standard für Museumsdokumentationen. Es existiert ein SPECTRUM XML-Schema, das für die Beschreibung von Museumsobjekten eingesetzt wird.

Dokumentation:

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 ist ein Standard für «historische Umgebungen» (historic environments) und legt fest, welche Informationen aufgezeichnet werden müssen, damit der Informationsaustausch und der Langzeiterhalt effektiv stattfinden können. MIDAS berücksichtigt andere Standards wie z. B. SPECTRUM.

Dokumentation:
<http://www.ica.org/>

10.3.1.10 INTERNATIONAL STANDARD FOR ARCHIVAL DESCRIPTION GENERAL (ISAD[G])

ISAD(G) ist ein internationaler Standard für die archivische Verzeichnung, der vor allem im Bereich der Schriftgutverwaltung (records management) eingesetzt wird. Der Standard gibt vor, welche Informationen einer Sammlung aufgezeichnet werden sollen, schreibt aber keine explizite Datenstruktur vor. ISAD(G) ist hierarchisch aufgebaut (Archiv, Bestand, Dossier (Akte), Dokument).

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

10.3.1.11 ENCODED ARCHIVAL DESCRIPTION (EAD)

Die Encoded Archival Description Document Type Definition (EAD DTD) ist ein in SGML (Standard Generalized Markup Language, Vorläufer von HTML) oder XML codierter Standard für die Beschreibung archivischer Findmittel. Der hierarchische Aufbau ermöglicht Beschreibungen auf unterschiedlichen Ebenen und ist für Kombinationen von digitalen Objekten gedacht (Sammlungen, Gruppierungen von Archivalien, Einzelstücke, vgl. ISAD(G)). EAD kann auch für die Beschreibung von Museumsobjekten eingesetzt werden.

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

10.3.1.12 DESCRIBING ARCHIVES: A CONTENT STANDARD (DACS)

DACS ist ein US-amerikanischer Standard für die archivische Verzeichnung und basiert auf ISAD(G). Der Standard zeigt auf, wie Informationen über Archivressourcen in Schemata wie EAD oder MARC21 überführt werden können.

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

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

TEI definiert eine Reihe von Datenelementen für die Auszeichnung von Texten in SGML oder XML. TEI-Lite ist eine Untermenge von TEI DTD und wurde bisher für die elektronische Transkription von Büchern und anderen Textmaterialien eingesetzt. Jedes TEI-Dokument hat einen Header (Kopfdaten) mit Metadaten, die Quelle, Entstehung und die elektronische Transkription dokumentieren.

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

Vgl. auch NISO-MIX:
<http://www.loc.gov/standards/mix/>

10.3.1.14 NISO (NISO TECHNICAL METADATA FOR DIGITAL STILL IMAGES)

Der Standard Z39.87 der NISO ist ein technischer Metadatenstandard für Raster-Bilder (pixel-basiert). Der Standard ist gleichzeitig auch ein Data Dictionary.

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

10.3.1.15 METADATA ENCODING AND TRANSMISSION STANDARD (METS)

METS beschreibt digitale Objekte anhand eines XML-Schemas. METS kann für den Datenaustausch, für die Präsentation im Web und für die digitale Archivierung eingesetzt werden.

Die METS-Spezifikation für ein Dokument besteht aus sieben Hauptabschnitten: Kopfteil (METS-Header), Erschließungsangaben (Descriptive Metadata), Verwaltungsangaben (Administrative Metadata), Dateiabschnitt (File Section), Strukturbeschreibung (Structural Map), Strukturverknüpfungen (Structural Links) und Verhalten (Behavior).

Wegen des komplexen XML-Schemas gestaltet sich die manuelle Codierung schwierig. Deshalb werden METS-Daten idealerweise aus Datenbanken generiert.

Dokumentation:
<http://www.oclc.org/research/projects/pmwg/>

10.3.1.16 PRESERVATION METADATA IMPLEMENTATION STRATEGIES (PREMIS)

PREMIS ist weniger auf die Definition eines expliziten Metadatenstandards ausgelegt. Die Arbeitsgruppe hat ein Data Dictionary ausgearbeitet, das eine «implementierbare Menge von Kern-Metadatenelementen» für die langfristige Aufbewahrung von digitalen Objekten mit breiten Anwendungsmöglichkeiten definiert. Das PREMIS-Datenmodell beinhaltet sieben Hauptentitäten: Objekt (object), intellektuelle Einheit (intellectual entity), Ereignis (event), Agent (agent), Recht (right), Beziehung (relationship) und semantische Einheit (semantic unit).

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

10.3.1.17 LEARNING OBJECT METADATA (LOM) UND LOM CORE

IEEE LOM ist ein in XML codiertes Metadatenschema für die Beschreibung von Lernressourcen. LOM Core ist eine für Lernkontexte des Vereinigten Königreichs (UK) optimierte Version von LOM.

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

10.3.1.18 OPEN ARCHIVES INITIATIVE PROTOCOL FOR METADATA HARVESTING (OAI_PMH)

OAI ist eine Arbeitsgemeinschaft, die sich über das OAI Protokoll mit dem Sammeln (harvesting) und der Weitergabe (dissemination) von Metadaten befasst und die Interoperabilität zwischen den verschiedenen Systemen unterstützt.

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

10.3.1.19 EXCHANGEABLE IMAGE FILE FORMAT (EXIF)

Exif ist ein von der Firma JEITA (Japan Electronics and Information Technology Industries Association) entwickelter technischer Metadatenstandard für Bild-Dateien. Die Daten werden wie bei IPTC direkt in die Datei geschrieben (TIFF und JPEG).

Dokumentation:
<http://www.iptc.org/pages/index.php>

10.3.1.20 INTERNATIONAL PRESS AND TELECOMMUNICATIONS COUNCIL (IPTC)

Von der IPTC entwickeltes Metadatenchema für beschreibende Metadaten, die direkt in die Bild-Datei abgelegt werden können (TIFF und JPEG). Das Beschreiben von IPTC-Feldern wird von den meisten Bildverarbeitungsprogrammen unterstützt. Eine XML-basierte Variante, IPTC G2, ist in Arbeit.

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

10.3.1.21 EXTENSIBLE MARKUP LANGUAGE (XMP)

XML-basierter Metadatenstandard der Firma Adobe, der zunehmend von Drittanbietern unterstützt wird. Ursprünglich als Metadatenstandard für die eigene Softwareproduktlinie entwickelt, unterstützt XMP auch andere Standards und erlaubt unter anderem auch das Schreiben von IPTC-Informationen in andere Bild-Dateien als TIFF und JPEG (JPEG2000, PNG, DNG). XMP schreibt und verwaltet nicht nur IPTC-Daten. Die Informationen lassen sich über die einzelnen Anwendungen hinaus verwalten und anzeigen.

10.4 KONTROLLIERTE VOKABULARE UND THESAURI

Kontrollierte Vokabulare und Thesauri unterstützen die einheitliche und konsistente Beschreibung (Beschlagwortung) digitaler Objekte. Weitere Vorteile sind eine kontrollierte Synonymie, gezielte Such- (Filter) und Strukturierungsmöglichkeiten. Damit dies funktioniert, wird Disziplin bei der Beschlagwortung abverlangt. Je nach Gebiet muss ein Spezialist mit dieser Aufgabe betraut werden. Nachteile ergeben sich grundsätzlich bei zwingenden Erweiterungen des Vokabulars.

Im deutschsprachigen Raum gibt es im Gegensatz zu den USA wenige kontrollierte Vokabulare. In den meisten Bibliotheken wird die deutsche Schlagwortnormdatei (SWD) verwendet.

Es ist jederzeit möglich, anwendungs- und fachspezifische Vokabulare zu entwickeln. Solange diese Vokabulare «proprietär» genutzt werden, entstehen keine Probleme. Sobald aber Daten ausgetauscht werden sollen, kann die Interoperabilität mit anderen Systemen nur mit zusätzlichem Aufwand und Mehrkosten ermöglicht werden. Deshalb sollten untereinander operierende Institutionen ein – bis auf sinnvolle Ausnahmen – gleiches Vokabular nutzen.

Im angelsächsischen Raum gibt es eine grosse Anzahl kontrollierter Vokabulare, die für den deutschen Sprachraum und speziell für die Schweiz von geringerer Relevanz sind. Dennoch liefern folgende Quellen nützliche Informationen:

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

10.4.1.1 ART & ARCHITECTURE THESAURUS® (AAT)

Strukturiertes Vokabular / Terminologie für Kunst, Architektur und verwandte Disziplinen.

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

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

Eher ein Metadatenschema, das Kategorien für Kunst, Architektur usw. beschreibt (vgl. 10.3.1.3).

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

10.4.1.3 GETTY THESAURUS OF GEOGRAPHIC NAMES (TGN)

Strukturiertes Vokabular, insbesondere für geografische Beschreibungen.

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

10.4.1.4 GRAPHIC MATERIALS

Regeln für die Katalogisierung von Drucken, Fotografien und anderem Bildmaterial.

www.ica.org

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

Beinhaltet 26 Elemente für die korrekte Identifikation archivspezifischer Materialien.

www.iconclass.nl

10.4.1.6 ICONCLASS

Ein Klassifikationssystem für die Ikonografieforschung, Bilddokumentation usw.

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

10.4.1.7 LIBRARY OF CONGRESS SUBJECT HEADINGS (LCSH)

Das grösste allgemeine Beschlagwortungsvokabular in englischer Sprache.

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

10.4.1.8 MOVING IMAGE MATERIALS: GENRE TERMS

Terminologiesammlung für die Beschlagwortung von Filmen.

<http://www.rarebookstore.net/cgi-bin/schuylers/5577.html?id=Jq6air9t>

10.4.1.9 SUBJECT INDEX FOR THE VISUAL ARTS

Strukturiertes Vokabular für den Zugriff auf Objekte im Print Room of the Victoria and Albert Museum in London.

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

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

10.4.1.10 THESAURUS FOR GRAPHIC MATERIALS (TGM)

Strukturiertes Vokabular mit Schlagworten für visuelle Materialien.

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

10.4.1.11 UNION LIST OF ARTIST NAMES® (ULAN)

Strukturiertes Vokabular mit Namen und anderen Informationen über Kunstschaffende.

11 ABKÜRZUNGSVERZEICHNIS

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
Exif	Exchangeable Image File Format
HTML	Hypertext Markup Language
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
JPEG2000	Neues Dateiformat der Joint Photographic Experts Group
JPG	Gängige Dateierweiterung von JPEG-Dateien
LOM	Learning Object Metadata
MARC	Machine Readable Cataloging (record)
MARC21	MARC Concise Formats-Familie
MARXML	MARC als XML-Schema
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)
PNG	Portable Network Graphics

PREMIS	PREservation Metadata Implementation Strategies
PSD	Photoshop Data Format
RDF	Resource Description Framework
RLG	Research Libraries Group
SEPIA	Safeguarding European Photographic Images for Access
SEPIADES	SEPIA Data Element Set
SGML	Standard Generalized Markup Language
SPECTRUM	The UK Museum Documentation Standard
TASI	Technical Advisory Service for Images
TEI	Text Encoding Initiative
TIFF	Tagged Image File Format
UNIMARC	MARC Format für den internationalen Austausch
VRA Core	Visual Resources Association Core
XML	eXtensible Markup Language
XMP	eXtensible Metadata Platform

NOTIZEN

KONTAKTADRESSEN

AUFTRAGGEBER

Bundesamt für Bevölkerungsschutz BABS
Fachbereich Kulturgüterschutz KGS
Rino Büchel, Reto Suter
Monbijoustrasse 51A
3003 Bern
www.kulturgueterschutz.ch
Tel.: +41 (0)31 325 15 37
Fax: +41 (0)31 324 87 89
E-Mail: reto.suter@babs.admin.ch

KONTAKTSTELLE FÜR FACHFRAGEN

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

AUTOREN DER VORLIEGENDEN GUIDELINES

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