

# Digitale Archivierung von fotografischen Sammlungen

–

## Ein Grundlagenbericht



**Datum:** April 2002

**erstellt für:** Bundesamt für Zivilschutz  
Sektion Kulturgüterschutz  
Herrn R. Büchel  
Monbijoustrasse 91  
**3003 Bern**

**von:** PD Dr. Rudolf Gschwind  
Dr. Lukas Rosenthaler

Inst. für Medienwissenschaften  
Universität Basel  
Klingelbergstrasse 80  
4056 Basel

Prof. Dr. Franziska Frey

College of Imaging Arts and Sciences  
Rochester Institute of Technology  
Rochester, NY

|   |    |
|---|----|
| 1. Einleitung.....  | 3  |
| 2. Richtlinien zur digitalen Langzeitarchivierung .....                 | 4  |
| 2.1. Einleitung.....  | 4  |
| 2.2. Prozedurale Kriterien.....   | 6  |
| 2.3. Technische Kriterien.....  | 10 |
| 3. Metadaten .....  | 17 |
| 3.1. Verschiedene Formen von Metadaten .....                            | 17 |
| 3.2. Grundsätze zur Archivierung von Metadaten .....                    | 17 |
| 3.3. Verschiedene Standards.....  | 17 |
| 3.4. Beschreibende Metadaten .....                                      | 18 |
| 3.4.1. Ordnungsdaten/Administrative Metadaten.....                      | 19 |
| 3.4.2. Technische Metadaten.....  | 19 |
| 3.5. Referenzen .....   | 24 |
| 4. Richtlinien zur Digitalisierung von Fotografien.....                 | 25 |
| 4.1. Einleitung.....  | 25 |
| 4.2. Digitalisierung.....   | 26 |
| 4.2.1. Auswählen der Hardware.....                                      | 26 |
| 4.2.2. Räumliche Auflösung.....   | 27 |
| 4.2.3. Tonwertumfang.....   | 27 |
| 4.2.4. Kalibrierung und Farbproduktion.....                             | 29 |
| 4.2.5. Dunkelstrom, Rauschen (Noise) und Empfindlichkeit.....           | 31 |
| 4.2.6. Geometrische Verzerrungen.....                                   | 32 |
| 4.3. Qualitätskontrolle .....   | 32 |
| 5. Anhang .....   | 34 |
| 5.1. Praktisches Beispiel einer digitalen Langzeitarchivierung.....     | 34 |
| 5.1.1. Digitalisierung und Qualitätskontrolle.....                      | 34 |
| 5.1.2. Migration.....   | 35 |
| 5.2. Workflow-Schema zum Digitalisieren fotografischer Materialien..... | 36 |

# 1. Einleitung

Der vorliegende Bericht beinhaltet Richtlinien zur sicheren Langzeitarchivierung von fotografischen Bildern als digitale Datensätze. Der Bericht besteht aus 3 Teilen:

## 1. **Richtlinien zur digitalen Langzeitarchivierung**

Dieser Teil beschreibt die Voraussetzungen, die ein Datenarchiv erfüllen muss, um digitale Bilddaten auch für kommende Generationen verfügbar zu halten.

## 2. **Metadaten**

Metadaten, d.h. Informationen über die Bilddaten, müssen einen integralen Bestandteil eines digitalen Bildarchivs bilden. In diesem Teil werden die gängigsten Formen sowie die aktuellen Standards im Bereich Metadaten beschrieben.

## 3. **Richtlinien zur Digitalisierung**

Um ein digitales Langzeit-Bildarchiv zu erstellen, müssen die Bilder in einer adäquaten Qualität digitalisiert werden. Die Einhaltung der hier beschriebenen Richtlinien garantiert, dass die digitalen Kopien der Originalfotografien in fast allen Anwendungen als „digitales Substitut“ dienen können.

Dem Bericht liegt der Grundsatz zugrunde, dass das digitale Substitut als vollwertiger Ersatz für alle Anwendungen dient, wo nur der visuelle Inhalt des Bildes von Interesse ist.

Im Bereich der Langzeitarchivierung muss das digitale Archiv mindestens die gleiche oder bessere Sicherheit der „traditionellen“, analogen Methode bieten. Auch ein traditionelles Archiv mit Sicherheitskopie auf Mikrofilm bietet keine 100% Sicherheit!

Das Thema Metadaten wird oft unterschätzt. Viele fotografische Sammlungen sind schlecht erschlossen, und die Metadaten – sofern überhaupt existierend – sind oft unsystematisch und nicht organisiert (z.B. nur handschriftliche Notizen auf der Rückseite der Fotografien). Der Aufwand, diese Metadaten in eine organisierte Form zu bringen, ist nicht zu unterschätzen, ebenso wenig der Aufwand, diese Daten in die digitale Form zu überführen.

Die Digitalisierung folgt wiederum dem Prinzip, dass die digitalen Kopien alle Anwendungen (oder noch mehr) bieten müssen, welche mit dem Originalbild möglich sind (Detailvergrößerung etc.). Ist dies erfüllt, so kann von einem „digitalen Faksimile“ gesprochen werden.

Die in diesem Bericht umfassend beschriebenen Richtlinien ermöglichen es, erfolgreich ein digitales Langzeitarchiv für Bilder zu implementieren.

## 2. Richtlinien zur digitalen Langzeitarchivierung

### 2.1. Einleitung

Im Bereich der Langzeitarchivierung analoger Daten (insbesondere Bilder fotografischer Natur) müssen vor allem zwei Risiken minimal gehalten werden: 1) die totale Zerstörung des Bildes, z.B. durch Feuer oder Wasser, und 2) muss der intrinsische Zerfall durch die natürliche Alterung des Mediums möglichst verlangsamt werden. Beide Risiken können durch eine sachgemäße Lagerung (Lagerort, Brandschutz, kontrolliertes Klima etc) optimiert werden. Konsequenterweise ist die Zugänglichkeit solchermaßen archivierter Daten sehr eingeschränkt, und der intrinsische Zerfall kann nicht etwa aufgehoben, sondern nur verlangsamt werden. Da jeder Kopiervorgang analoger Daten (z.B. von Fotografien, Film, Video,...) mit einer z.T. erheblichen Qualitätseinbusse einhergeht, *muss* das Original für die Nachwelt erhalten werden. Vielfach haben analoge Daten die Eigenschaft, dass sie ohne weitere technische Hilfsmittel direkt interpretiert werden können (z.B. Fotografie). Metadaten wie z.B. Bildherkunft, Autor, Objektbeschreibung etc. können oftmals auf dem Medium selbst festgehalten werden, z.B. in Form von Notizen auf der Rückseite einer Fotografie.

In Bezug auf Langzeitarchivierung haben digitale Daten gegenüber analogen Daten völlig andere Eigenschaften:

1. Digitale Daten lassen sich, das richtige Prozedere vorausgesetzt, beliebig oft kopieren. Original und Kopie sind identisch und ununterscheidbar, weshalb der Begriff „Original“ in der digitalen Domäne seine Bedeutung verliert.
2. Es gibt in der digitalen Domäne keinen kontinuierlichen Zerfall. Digital aufgezeichnete Information kann entweder vollständig (und richtig) gelesen werden, oder es treten Fehler auf, welche im Prinzip den ganzen betroffenen Datensatz wertlos machen. Um das Auftreten von Fehlern überhaupt erkennen zu können, müssen spezielle Algorithmen („Checksums“) eingeführt werden<sup>1</sup>.

Welche Mechanismen können nun zu einem Informationsverlust bei digitalen Daten führen? In der folgenden kurzen Übersicht wird speziell auf die Probleme der Langzeitarchivierung eingegangen, wo nicht nur das Objekt an sich zu verwahren ist, sondern auch die Information über das Objekt, die Metadaten<sup>2</sup>. Bei digitalen Archiven kommt noch eine weitere Schwierigkeit dazu: Die Datenträger selbst können nur mit technischen Hilfsmitteln gelesen und interpretiert werden. Ein Computertape sieht von aussen immer gleich aus, unabhängig davon ob es leer ist, ob Bilder, Texte oder andere Daten darauf gespeichert sind. Ein Schrank mit

---

<sup>1</sup> Bei heutigen Datenträgern ist die Berechnung und Prüfung von „Checksums“ meistens in die Hardware integriert. Im Falle von Lesefehlern wird oft der ganze Datensatz als unleserlich bezeichnet und übersprungen bzw. der Lesevorgang ganz abgebrochen. Insofern gibt es bei digitalen Daten nur 2 Möglichkeiten: der Datensatz kann gelesen werden und ist deshalb auch „korrekt“, oder der Datensatz ist unleserlich und verloren.

<sup>2</sup> Metadaten sind definiert als Informationen über den Kontext eines Objektes. Bei Fotografien können es z.B. eine textuelle Beschreibung, Namen der abgebildeten Personen, Ort und Zeit, Photograph, aber auch technische Informationen wie Filmtyp, Belichtungszeit etc. sein. Weitere Informationen sind unter dem Kapitel „Metadaten“ enthalten.

10`000 CD-ROMs, welche nicht angeschrieben sind, ist nahezu wertlos, wenn nicht weitere Informationen vorhanden sind. Daraus folgt, dass bei unzweckmässigen Archivierungsstrategien digitaler Daten auf mehreren Ebenen ein Informations-verlust stattfinden kann. Ist nur auf einer dieser Ebenen ein Informationsverlust aufgetreten, so wird der Nutzen und Wert eines digitalen Archivs erheblich gemindert oder vollständig zerstört.

Grundsätzlich können die folgenden 6 Ebenen unterschieden werden, auf denen ein Informationsverlust auftreten kann:

1. **Ordnungsdaten/administrative Metadaten:** Die gesuchte Information wird nicht mehr gefunden, weil keine Ordnungsdaten vorhanden oder verfügbar sind. Unter Ordnungsdaten verstehen wir all jene Informationen, welche dazu dienen, die digital abgespeicherten Daten zu finden und zu lesen, insbesondere Dokumentationen über Lagerort, verwendete Formate und Verfahren, technische Angaben über Datenträger etc. In einer gewissen Weise sind die Ordnungsdaten die Metadaten des Archivs (und nicht diejenigen des *Archivinhalts*).
2. **(beschreibende) Metadaten:** Die Objektdaten sind vorhanden und lesbar, es fehlen aber die Beschreibungen der Objekte, oder sie sind fehlerhaft. (z.B.: es sind lesbare Bilder auf einem Datenträger, aber der weitere Zusammenhang der Bilder wie Ort, Zeitpunkt, abgebildete Personen und Objekte etc. ist nicht bekannt)
3. **Dateiformate:** Die Dateien können zwar als Bitinformation gelesen werden, aber das Dateiformat ist unbekannt und kann nicht mehr interpretiert werden, z.B. wegen fehlender Software.
4. **Datenträger-Formate:** Der Datenträger kann nicht gelesen werden, weil die Formatierung des Datenträgers nicht bekannt oder nicht mehr unterstützt wird. Unter der Formatierung eines Datenträgers wird die Art und Weise verstanden, wie die Daten auf dem Datenträger gespeichert werden. Es ist durchaus möglich, dass für einen Typ von Datenträgern mehrere Formate bestehen, wie die Daten geschrieben/gelesen werden. So kann ein Magnetband des Typs DAT sowohl mit NT-Backup, mit „tar“ oder als „ANSI-labeled tape“ beschrieben werden. Alle diese Formate sind untereinander inkompatibel, d.h. ein mit NT-Backup beschriebenes Magnetband kann nicht als „tar“ gelesen werden. Ein weiteres Beispiel bilden die verschiedenen Formate, mit denen CD-R's beschrieben werden können: ISO9660, Joliet, UDF, DirectCD, UDF, Mac etc. Diese Formate sind untereinander inkompatibel und können z.T. nur auf bestimmter Hardware oder mit spezieller Software gelesen werden.
5. **Lesegeräte:** Der Datenträger kann nicht gelesen werden, weil die entsprechenden Lesegeräte nicht mehr verfügbar und unterstützt sind.
6. **Datenträger/Medium:** Der Datenträger – das Medium – ist nicht mehr lesbar infolge Beschädigung durch Alterung, Fehlbedienung etc.

Auf allen diesen Ebenen müssen entsprechende Vorkehrungen getroffen werden, um die digitale Information über lange Zeiten bewahren zu können und verfügbar zu halten.

## 2.2. Prozedurale Kriterien

Um die Langfristigkeit digitaler Daten zu garantieren, müssen gewisse grundsätzliche Vorgehensweisen, Regeln und Strategien eingehalten werden. Ein digitales Langzeitarchiv zeichnet sich insbesondere dadurch aus, dass mit den zu archivierenden Daten laufend gearbeitet wird. Laufende Kontrollen und permanentes Umkopieren/Umformatieren garantieren erst eine (im Prinzip) unbeschränkte Lebensdauer der Daten. Diese Arbeitsprozesse müssen jedoch sehr strengen Qualitätsmerkmalen genügen, und der gesamte Archivierungsprozess muss gewisse Grundsätze berücksichtigen:

**Frage:** Wurde ein Langzeit-Archivierungskonzept erstellt?

- **Redundanz:**

Redundanz auf allen Ebenen, d.h. die Wahrscheinlichkeit eines totalen Informationsverlusts, muss sich asymptotisch 0 nähern. Da sich digitale Daten einfach duplizieren oder „klonen“ lassen, kann auf allen Risikoebenen das Risiko eines Datenverlustes durch Redundanz stark verkleinert werden.

- a. *Redundanz auf Ebene Datenträger*

Mehrfache identische Datenträger: Die Daten werden nicht nur einmal, sondern identisch auf mehreren Datenträgern abgespeichert. Ist die Wahrscheinlichkeit, dass ein einzelner Datenträger nicht mehr gelesen werden kann, 1/1000 (einer auf Tausend), so ist bei 3 identischen Datenträgern (dreifache Redundanz) die Wahrscheinlichkeit, dass alle 3 Datenträger nicht mehr gelesen werden können 1/1000000000 (einmal auf eine Milliarde Fälle).

**Frage:** Wie viele Kopien eines Datenträgers werden hergestellt und „bewirtschaftet“ (Qualitätskontrolle, Migration etc.)?

- b. *Geographische Verteilung*

Die redundanten Datenträger sollten an geografisch verschiedenen Orten aufbewahrt werden, um das Risiko eines Totalverlustes durch ein katastrophales Ereignis (Feuer, Erdbeben, kriegerische Ereignisse etc.) zu eliminieren.

**Frage:** Werden die redundanten Kopien

- *im gleichen Raum*
    - *im gleichen Gebäude*
    - *am gleichen geographischen Ort*
    - *landes- oder weltweit verteilt aufbewahrt?*

- **Migration**

Da eine regelmässige Migration der Daten, bedingt durch Änderungen der Technologie oder Alterung der Medien, unumgänglich ist, muss eine langfristige Migrationsstrategie implementiert werden. Diese Migrationsstrategie muss folgende Eigenschaften aufweisen:

a. *Migrationszeitpunkt:*

Die Daten müssen früh genug migriert werden, bevor ein Datenverlust durch Alterung des Mediums oder durch Technologiewandel auftreten kann.

**Frage:** *Sind Migrationen eingeplant und organisatorisch und technologisch vorbereitet?*

b. *Periodische Überprüfung:*

Die Datenträger und die darauf abgespeicherten Daten müssen periodisch überprüft werden, um z.B. frühzeitiges Altern durch Materialfehler etc. rechtzeitig zu erkennen<sup>3</sup>.

**Frage:** *Werden die Datenträger periodisch prüfgelesen und auf Konsistenz getestet?*

c. *NULL Fehlertoleranz:*

Alle Kopiervorgänge müssen mit Null Fehlertoleranz durchgeführt werden, um jedem Informationsverlust vorzubeugen. Dies kann erreicht werden, indem jede Kopie sofort nach dem Schreiben mit dem „Original“ verglichen wird.

**Frage:** *Wird jeder Schreibvorgang auf Korrektheit überprüft (Prüflesen unmittelbar nach Schreiben von Daten mit Vergleich „Original“ mit „Kopie“)?*

d. *Rollende Migration:*

Der stete Wandel der Speichertechnologie bedingt, dass die Daten periodisch auf eine neue Technologie migriert werden müssen, da z.B. die Wartung der „alten“ Geräte zunehmend teurer oder gar unmöglich wird (fehlende Ersatzteile etc.). Zur Erhöhung der Redundanz in Bezug auf Technologie sollte der Migrationsprozess von einer Generation von Geräten auf die nächste **rollend** erfolgen: Eine neue Technologie muss eingeführt werden, bevor die alte Technologie obsolet geworden ist. In anderen Worten: die „aktive“ Generation von Speichermedien verteilt sich auf mindestens zwei Technologien, wobei eine davon „altbewährt“ sein sollte.

**Frage:** *Sind die Migrationen rollend geplant, so dass die Medien der „alten“ Generation noch einiger Zeit nach der Migration auf ein neues Medium unterstützt und aktiv sind?*

<sup>3</sup> Eine solche periodische Überprüfung kann z.B. beinhalten, dass mehrere (redundant vorhandene) Datenträger mit identischen Daten miteinander verglichen werden. Dabei dürften keine Differenzen auftreten. Im Falle von Differenzen wird eine weitere Kopie als Entscheidungshilfe gelesen. Daraus folgt, dass immer mindestens drei identische Kopien vorhanden sein müssen.

- **Medienqualität**

In der analogen Fotografie (um ein Beispiel zu nennen) haben sich Verfahren durchgesetzt, die eine gute Aussage über die Langzeitstabilität eines bestimmten Materials erlauben (Schnellalterung, Arrhenius-Plots, Essigsäurebestimmung etc.). Bei digitalen Datenträgern, die für die Langzeitarchivierung bestimmt sind, müssen derartige Verfahren, soweit vorhanden, ebenfalls angewendet werden.

- a. *Prüflesen*

Bei allen heute bekannten digitalen Aufzeichnungsverfahren sind Aufzeichnungsfehler in einem kleinen Ausmass unvermeidbar, z.B. bedingt durch nicht perfekte Medien (Materialfehler). Diese intrinsischen Aufzeichnungsfehler werden aufgefangen, indem Fehler-Korrekturverfahren („error-correction“ der „recoverable errors“) angewandt werden<sup>4</sup>. Somit kann trotzdem garantiert werden, dass die Daten korrekt wiedergegeben werden. Erlaubt ein Aufzeichnungsgerät, die Anzahl der erfolgreichen Fehlerkorrekturen zu bestimmen, so erhält man damit ein Mass für die Qualität der Kombination Medium plus Lesegerät. Ein regelmässiges Prüflesen erlaubt festzustellen, ob die Anzahl Fehler zugenommen hat, und ist somit ein guter Indikator, um schadhafte (vorschnell alternde) Medien frühzeitig zu erkennen.

**Frage:** Werden beim periodischen Prüflesen die Anzahl der korrigierten Fehler erfasst und ausgewertet?

- b. *Medienchecks*

Die Qualität von Datenträgern variiert von Hersteller zu Hersteller, und kann auch beim selben Hersteller in der Zeit erheblich variieren. Deshalb sollte im Prinzip jede Charge von Medien stichprobenartig auf die Qualität überprüft werden, um rechtzeitig herstellungsbedingte Qualitätsmängel (Materialfehler) zu erkennen. Bei einigen Medien (z.B. CD-R) ist eine optimale Abstimmung von Aufzeichnungsgerät und Medium notwendig, um eine optimale Qualität und damit Langzeitsicherheit zu erhalten. Im Falle von CD-R sind gute Ergebnisse (d.h. CD's mit sehr wenig Fehlern) nur durch eine optimal aufeinander abgestimmte Kombination von CD-Brenner – Rohling – Schreibgeschwindigkeit erreichbar. Da mit normaler Aufzeichnungshardware kaum Aussagen über die Güte gemacht werden können, muss spezielle Prüfhardware eingesetzt werden, um die optimale Kombination herauszufinden. Dieser Test sollte für jede neue Charge von Medien wiederholt werden.

---

<sup>4</sup> Diese Fehlerkorrekturverfahren beruhen auf ausgeklügelten mathematischen Verfahren, welche einerseits das Erkennen von Fehlern und deren Korrektur in einem Schritt erlauben. Diese Verfahren garantieren bis zu einer gewissen Fehlerhäufigkeit, dass die digitalen Daten im streng mathematischen Sinn absolut korrekt wiedergegeben werden. Ein weit verbreitetes Verfahren beruht auf dem „cyclic redundancy check“ und wird als CRC-Verfahren bezeichnet.

**Frage:** Werden die Datenträger systematisch auf Defekte, Materialfehler etc. untersucht, bevor sie in den Archivierungszyklus gelangen?  
Falls ja, wird dies

- individuell für jeden Datenträger
- stichprobenartig für jede Charge
- stichprobenartig für jeden Datenträgertyp (Hersteller, Lieferant)
- allgemein nur stichprobenartig ohne besonderes System durchgeführt

c. *Optimale Aufbewahrungsbedingungen der Datenträger (Umwelt, Klima)*

Datenträger sind relativ kurzlebig<sup>5</sup>. Um die Lebensdauer nicht weiter zu verkürzen, müssen die Medien nach Vorschriften gelagert werden. Bei Magnetbändern wird eine Lagertemperatur um 15°C (Schwankung  $\pm 2^\circ\text{C}$  bei einer rel. Feuchtigkeit von 20-40% (Schwankung  $\pm 5\%$ ) vorgeschrieben (SMTPE, RE 103 oder ANSI/AES).

**Frage:** Entsprechen die Lagerbedingungen den Normen bzw. den optimalen Klimabedingungen für den jeweiligen Datenträger?  
**Frage:** Sind Einrichtungen zur Sicherung gegen Elementarschaden bzw. Einbruch und böswillige Zerstörung vorhanden?

1. **Risiken bei Handling**

Ein grosses Risiko für Datenverlust ist der menschliche Faktor. Ein Datenträger kann fallen gelassen werden, er kann mit Kaffee überschüttet oder schlichtweg verlegt werden. Auch Fehlmanipulationen, welche zum Löschen eines Datenträgers geführt haben, sollen schon vorgekommen sein... Um das Risiko menschlicher Irrtümer so klein wie möglich zu halten, müssen alle Schritte systematisch geplant, sorgfältig und langsam durchgeführt und dokumentiert werden. Selbst ein kleiner Fehler, wie eine falsche Beschriftung kann zu erheblichem Datenverlust führen. Deshalb sollten die Verarbeitungsschritte, welche Datenträger manipulieren, möglichst automatisiert werden, und der ganze Prozess ist einer strikten Qualitätskontrolle zu unterziehen.

**Frage:** Existiert ein System der Qualitätskontrolle im Bereich Handling der Datenträger?  
**Frage:** Sind die Prozesse (Archivierung, Prüfllesen, Migration etc.) weitgehend automatisiert?

<sup>5</sup> Bei guten Magnetbänder rechnet man mit 15 – 30 Jahren, CD-R's erreichen (schätzungsweise) im optimalsten Falle 50 – 100 Jahre. Das ist länger als die eigentliche Systemlebensdauer von Massenspeichern, die beim aktuellen rasanten Wechsel in der EDV-Industrie mit maximal 7 Jahren veranschlagt werden darf.

## 2.3. Technische Kriterien

Um die Langfristigkeit digitaler Daten garantieren zu können, müssen auf jeder der eingangs erwähnten Ebenen gewisse Kriterien erfüllt werden, um das Risiko eines Datenverlustes minimal zu halten. Im Folgenden sollen diese Kriterien detailliert erläutert und begründet werden:

### 1. Ordnungsdaten/administrative Metadaten

Die Ordnungsdaten dienen dazu, die Datenträger und den Zugriff darauf zu organisieren. Im einfachsten Fall bestehen Sie aus einer Liste, in der festgehalten ist, auf welchen Datenträgern sich die zu archivierenden Objekte befinden. Ordnungsdaten sind unerlässlich: einerseits, um die gewünschten Daten zu finden, und andererseits, um eine rechtzeitige und systematische Migration durchzuführen.

#### a. *Die Datenträger sind sinnvoll und auch für den Menschen lesbar (human and machine readable) zu bezeichnen.*

Begründung: Es hat sich gezeigt, dass auch bei bester Organisation der Arbeitsabläufe Irrtümer oder Fehler auftreten (z.B. falsch geleitete Postsendung etc.). In einem solchen Falle ist es von hohem Nutzen, wenn durch die Beschriftung eines Datenträgers direkt auf dessen Inhalt geschlossen werden kann. So sagt z.B. eine Nummer wie 467812 sehr wenig aus, während eine Beschriftung wie z.B. „Kunstmuseum Basel, Sammlung Sarasin, Bilder 1-250, 12.03.1999“ wesentlich mehr aussagt. Falls zusätzlich noch ein maschinenlesbares Label angebracht wird (z.B. Barcode), so kann dies die Effizienz erheblich erhöhen (die zur Interpretation notwendige Infrastruktur wie Barcode-Leser, Datenbank muss allerdings vorhanden sein).

**Frage:** Werden die Datenträger sinnvoll und lesbar beschriftet?

#### b. *„Buchführung“ der Daten und Datenträger (Version, Standort, Format, Datum...).*

Im Rahmen der Arbeitsabläufe sollte genau über alle Vorgänge und speziell über alle Bewegungen der Datenträger Buch geführt werden. Damit wird einerseits das Risiko eines Verlusts minimiert, und andererseits kann bei Problemen (z.B. Leseprobleme) durch eine genaue Analyse der „Geschichte“ eines Datenträgers Rückschlüsse auf die Ursache gezogen werden. Damit kann weiterer Schaden vermieden werden.

**Frage:** Gibt es eine entsprechende Buchführung bzw. Qualitätskontrolle (z.B. nach ISO 9001)?

#### c. *Ordnungsdaten zusätzlich in „analoger“ Form (z.B. Papier)*

Die Ordnungsdaten bilden die Basis der digitalen Langzeitarchivierung, ohne die ein digitales Archiv zu einer nutzlosen Ansammlung von Datenträgern wird. Deshalb sind die Ordnungsdaten, welche im Vergleich zum eigentlichen Archivinhalt

sehr kompakt sind, ebenfalls mit hoher Redundanz abzuspeichern. Um eine Interpretation der Ordnungsdaten – welche meist in Form von Text (Beschreibungen, Listen, Software-Sourcecode etc.) vorliegen – ohne technische Hilfsmittel zu erlauben, sollten die Ordnungsdaten zusätzlich auch in analoger Form, z.B. als Papierausdruck, vorliegen.

**Frage: Wie werden die Ordnungsdaten archiviert?**

- **Beschreibende Metadaten**

Die Metadaten eines Objektes (z.B. Bildbeschreibung, Aufnahmedatum, Photograph, Copyright etc.) werden zwar meist getrennt vom Objekt gespeichert, bilden aber eigentlich einen integralen Bestandteil der Objektinformation, da ein Objekt ohne die dazugehörigen Metadaten meist nur wenig Aussagekraft aufweist.

- a. *Verbindung Metadaten/Objekt*

Da die Metadaten einen integralen Bestandteil der Objektdaten darstellen, sollten sie *zusätzlich* zu komplexeren Speichermethoden wie z.B. relationalen Datenbanken auch direkt mit den Objektdaten verbunden werden, um das Risiko eines Verlusts der Verbindung der Metadaten mit einem dazugehörigen Objekt zu minimieren. Im Falle von Bildern gibt es zwei grundsätzliche Möglichkeiten: 1) die Metadaten werden im „Header“ (Kopf) der Datei abgespeichert<sup>6</sup>. 2) Die Metadaten werden direkt zum Bild in der Form von bildlicher Information hinzugefügt. So können z.B. die Metadaten als Texttafel beim Digitalisierungsprozess mit eingescannt werden (auch nachträglich). Damit sind die Metadaten ebenfalls unmittelbar mit dem zu archivierenden Objekt verbunden. Oft kann der Dateiname selbst verwendet werden, um ein Objekt für den Menschen aussagekräftig zu kennzeichnen.

**Frage: Werden (die „wichtigsten“) Metadaten der Objekte mit dem Objekt selbst archiviert?**

**Frage: Werden sinnvolle und aussagekräftige Dateinamen verwendet?**

- b. *Einfachheit, flaches Format*

Metadaten sind oft hoch strukturiert und werden meistens in der Form von relationalen Datenbanken organisiert. Da der Zugriff auf die Datenbanken komplexe Softwaresysteme erfordert, welche ebenfalls dem rapiden Wandel der Technologie unterworfen sind, sollte die Langzeitarchivierung der Metadaten in einer einfach strukturierten Form, z.B. einfache Listen oder Tabellen, erfolgen. Die so archivierten Metadaten können bei adäquater Ausführung jederzeit wieder in eine strukturierte Form überführt werden, welche für einen effizienten Zugriff (Suche) notwendig ist.

---

<sup>6</sup> Das TIFF-Format, ein weit verbreitetes Bildformat, erlaubt es, im Header eines Bildes Zusatzinformation wie z.B. eine Bildbeschreibung, den Autor, Copyright etc. als ASCII-Text abzuspeichern. Damit werden die Metadaten zu einem direkten Bestandteil des digitalen Bildes.

**Frage:** Sind die Metadaten in einer flachen Struktur archiviert?

c. *Redundanz der Metadaten, Verbindung mit Objekt*

Die Metadaten selbst sollten den gleichen Sicherheitsaspekten unterstehen wie die eigentlichen Objektdaten. Das heisst, es gelten die gleichen Anforderungen an Redundanz und geographische Verteilung wie für die eigentlichen Objektdaten.

**Frage:** Werden die Metadaten mit der gleichen Redundanz wie die Objektdaten archiviert?

d. *Checksummen*

Als Metadaten können auch „Checksummen“ der digitalen Files betrachtet werden. Diese können mit geeigneten Programmen bestimmt werden (? , ?) und müssen zusammen mit den eigentlichen Datenfiles mitgespeichert werden. Beim Prüfllesen der Daten kann so einfach allfällige Lese-Fehler festgestellt werden.

**Frage:** Sind im Archivierungskonzept Checksummen vorgesehen?

• **Dateiformat**

Für die Langzeitarchivierung geeignete Dateiformate sollten selbst eine möglichst lange Lebensdauer aufweisen. Dies kann erreicht werden, wenn die folgenden Kriterien möglichst gut erfüllt werden:

a. *Offene Definition*

Das zur Archivierung verwendete Dateiformat muss vollständig dokumentiert und offengelegt sein. Im Prinzip muss es möglich sein, dass ein Programmierer in der Lage ist, anhand der Dokumentation ein Programm zu entwickeln, welches die Daten lesen und richtig interpretieren kann. Nur so kann gewährleistet werden, dass die Daten unter allen Umständen in der (näheren) Zukunft noch interpretiert werden können. Für viele offen gelegte Standardformate wie das TIFF-Format existieren open-source<sup>7</sup> Programme, um das Format Lesen und Schreiben zu können. Diese Formate erfüllen die Forderung nach einer offenen Definition in idealer Weise.

b. *Verbreitung*

Das gewählte Dateiformat sollte eine möglichst grosse Verbreitung aufweisen. Damit ist einerseits sehr wahrscheinlich, dass das Dateiformat über lange Zeit aktuell bleibt. Eine grosse Verbreitung erhöht auch die Wahrscheinlichkeit, dass mehrere kommerzielle und open-source Anbieter dieses Dateiformat unterstützen.

c. *Flexibilität*

Das Dateiformat sollte entweder erweiterbar sein oder eine hohe Flexibilität aufweisen, um Kalibrierungsdaten, Objektdaten (wie z.B.

<sup>7</sup> „open-source“-Programme sind Softwarepakete, deren Programmtext („source-code“) öffentlich zugänglich ist, und deren Lizenzierung die Verwendung und Änderungen/Anpassungen Allen frei stellen.

die absolute Grösse des Originals in Millimeter) und Metadaten mit abzuspeichern.

d. *Fehlertoleranz*

Obwohl durch ein sorgfältiges Archivierungsprozedere das Auftreten von Datenfehlern äusserst klein ist, sollte das gewählte Datenformat gegenüber kleinen Fehlern (z.B. ein Bit falsch) möglichst tolerant sein. Im Bereich Bilddaten heisst dies zum Beispiel, dass ein kleiner Fehler in den Daten auch nur eine kleine visuelle Auswirkung hat. Bitfehler können je nach Format sehr unterschiedliche Auswirkungen haben. Im untenstehenden Beispiel ist im rechten Bild (als JPEG Format gespeichert) ein Bit falsch gesetzt worden, mit massiven Auswirkungen (grüner Balken). Im Allgemeinen kann davon ausgegangen werden, dass jegliche Form von Datenkompression die Auswirkungen eines Bitfehlers stark erhöhen!



Das gleiche Experiment, bei Bilddaten im unkomprimierten TIFF-Format ergibt eine visuell kaum sichtbare Veränderung.

**Frage:** *Welches Dateiformat wird verwendet?*

| Format            | Offenheit | Verbreitung | Flexibilität | Fehlertoleranz | Wertung |
|-------------------|-----------|-------------|--------------|----------------|---------|
| TIFF (raw)        | ++        | ++          | ++           | ++             | ++      |
| TIFF (LZW-kompr.) | ++        | ++          | ++           | -              | -       |
| PSD (Photoshop)   | --        | +           | ++           | 0              | 0       |
| JPEG              | ++        | ++          | ++           | --             | -       |
| BMP               | +         | +           | 0            | +              | +       |
| PNG               | ++        | +           | ++           | 0              | +       |
| PhotoCD           | -         | 0           | +            | 0              | 0       |
| GIF               | ++        | ++          | +            | 0              | 0       |
| Acrobat           | --        | +           | ++           | 0              | -       |

- **Format Datenträger<sup>8</sup>**

- a. *Offenheit*

Im Bereich von Formatierung von Datenträgern existieren leider viele proprietäre Formate, z.B. von Backup-Systemen. Für die Langzeitarchivierung muss ein Datenträgerformat gewählt werden, welches entweder nicht proprietär oder einen akzeptierten Industriestandard darstellt, welcher von mehreren Herstellern unterstützt wird. Als gutes Beispiel eines nicht proprietären Formats gilt das „tar“-Format, welches ursprünglich aus der Unix/Linux-Welt stammt, aber von allen Rechnerarchitekturen und Betriebssystemen unterstützt wird<sup>9</sup>. Ein proprietäres, aber weit verbreitetes Format stellt das NT-Backuptape dar. Weitere weit verbreitete Formate sind z.B. ISO9660 (CD-R), DOS-Format (Floppy), ANSI-labeled Tape etc.

**Frage:** Welche Formatierung des Datenträgers wird verwendet:

- Standardformat offen?
- Standardformat proprietär? (Unterstützung durch eine grosse Anzahl von Herstellern)
- Proprietäres Format? (nur ein Hersteller)

- **Hardware-System (Schreib- und Lesegeräte)**

Zu dem Hardware-System zählt das physische Format des Datenträgers (Typ, z.B. DAT, DLT, CD-R) mit den dazugehörigen Schreib- und Lesegeräten.

- a. *Verbreitung*

Das Speichersystem sollte eine möglichst grosse Verbreitung aufweisen und von mehreren Herstellern unterstützt werden. Dies ist insbesondere wichtig, da in der Informatikbranche auch grosse Hersteller innert kürzester Zeit vom Markt verschwinden können. Ist der einzige Hersteller eines bestimmten Hardware-Systems vom Markt verschwunden, so ist meistens auch der Support der Geräte in Frage gestellt und die so archivierten Daten verloren.

- b. *Verschleiss*

Das System sollte den Datenträger beim Lesen/Schreiben möglichst wenig beanspruchen. Da bei jedem Langzeitarchivierungskonzept die Daten während der Lebensdauer eines Datenträgers mehrmals prüfgelesen werden müssen, sollte die physikalische Beanspruchung des Datenträgers so klein wie möglich gehalten werden. Deshalb sind z.B. berührungsfreie optische Datenträger (z.B. magneto-optische Disks) unter diesem Aspekt von Vorteil. Im Bereich

<sup>8</sup> Eigentlich müsste im Bereich der Formate noch weiter differenziert werden: 1) Physisches Format des Datenträgers, 2) Format des Dateisystems (z.B. tar, FAT32, NTFS, UFS, ISO9660,...), 3) Dateiformat (TIFF, JPEG,...) und 4) Datenformat (Anzahl Bit, Bit- und Byte-Order, Format der Fließkommazahlen etc.). Diese Hierarchie wird noch dadurch erschwert, dass gewisse Formate (z.B. tar) sowohl das Format eines Dateisystems als selbst auch als Dateiformat (Datencontainer) vorliegen. Aus der praktischen Erfahrung ist die im Bericht vorgenommene Einteilung zweckdienlich und genügend. Im Bereich des Datenformats ist meist auf der Stufe Dateiformat vorgesorgt, dass die Daten immer richtig interpretiert werden (z.B. TIFF, JPEG).

<sup>9</sup> Bei Verwendung entsprechender Software

der Magnetbänder sind lineare Aufzeichnungsverfahren den „helical-scan“-Verfahren vorzuziehen<sup>10</sup>.

c. *Zuverlässigkeit*

Das System sollte eine möglichst hohe Zuverlässigkeit aufweisen. Ein Mass dafür ist die sog. „mean time between failure“ (MTF), welche aussagt, wie lange ein Gerät im Durchschnitt ohne Fehler läuft. Für die meisten Geräte existieren MTF-Angaben vom Hersteller.

d. *Austauschbarkeit*

Der Datenträger sollte ohne Probleme zwischen verschiedenen Aufzeichnungsgeräten ausgetauscht werden. Die Erfahrung hat gezeigt, dass z.B. helical-scan Verfahren in diesem Bereich weniger tolerant sind. Im Allgemeinen sollte jedes neue Aufzeichnungs-/Lesegerät auf die Austauschbarkeit mit anderen Geräten des gleichen Typs überprüft werden.

e. *Versionskompatibilität*

Bedingt durch die rasche technische Entwicklung der Speicherhardware entstehen oft Familien von Speichertechnologien, bei denen verschiedene Generationen des gleichen Typs über immer mehr Speichervolumen verfügen (z.B. DAT: DDS-1=2GB, DDS-2=8GB, DDS-3=24GB, DDS-4=40GB, DLT: DLT I – DLT IV). Da meistens nicht die Lebensdauer des Speichermediums an sich, sondern die Lebensdauer des Systems der beschränkende Faktor ist, sollte eine möglichst grosse Rückwärtskompatibilität gewährleistet werden (d.h. neue Geräte können auch alte Medien lesen).

f. *Fehlertoleranz*

Das Hardware-System sollte eine möglichst hohe Fehlertoleranz aufweisen, d.h. selbst bei Medienfehler durch Fehlererkennung und Fehlerkorrektur richtige Daten liefern. Zusätzlich sollte es möglich sein, die Anzahl der erfolgten Fehlerkorrekturen nach erfolgtem Lesen eines Mediums zu eruieren. Viele Hersteller geben an, nach wie vielen gelesenen Bits zum ersten Mal ein Fehler zu erwarten ist.

g. *Handling (Robustheit, Automatisierung)*

Das Hardwaresystem sollte mechanisch möglichst robust sein, um die Wahrscheinlichkeit der Zerstörung eines Datenträgers durch einen Defekt oder eine falsche Handhabung (!) möglichst klein zu halten (z.B. das bei Tonbandkassetten übliche „Verwickeln“ der Bänder). Im Weiteren sollte das System eine möglichst grosse Automatisierung zulassen, wie z.B. durch Tape-Roboter im Bereich DLT oder DAT.

**Frage: Welches Hardware-System wird verwendet?**

| System            | Verbr. | Verschleiss | Lebensd. | Austausch | Kompat. | Fehlertol. | Robust. |
|-------------------|--------|-------------|----------|-----------|---------|------------|---------|
| CD-R              | ++     | ++          | +        | +         | ++      | +          | +       |
| DVD <sup>11</sup> | +      | ++          | +        | -         | 0       | +          | +       |

<sup>10</sup> Im „helical-Scan“-Verfahren, welches ursprünglich für Videoaufzeichnung entwickelt wurde, wird das Magnetband um einen rasch rotierenden, etwas schräg stehenden Magnetkopf gewickelt. Dadurch werden die Daten in schrägen Spuren auf das Band aufgezeichnet.

<sup>11</sup> Unter DVD werden alle „DVD“ Medien zusammengefasst: DVDROM, DVD-R, DVD+R, DVD-RW, DVD+RW, DVD-RAM

|                |    |    |    |    |    |   |   |
|----------------|----|----|----|----|----|---|---|
| MO             | -- | ++ | ++ | +  | -  | - | + |
| Wechselplatte  | +  | 0  | 0  | +  | -  | + | 0 |
| DLT            | +  | ++ | ++ | ++ | ++ | + | + |
| Exabyte        | 0  | -  | -  | 0  | +  | 0 | 0 |
| DAT            | +  | -  | -  | 0  | ++ | 0 | + |
| AIT            | 0  | +  | +  | 0  | +  | 0 | 0 |
| LTO            | 0  | ++ | ++ | ++ | ++ | + | + |
| IBM/StorageTek | +  | ++ | ++ | ++ | ++ | + | + |

- **Medium**

- a. *Robustheit*

Das Medium selbst sollte möglichst robust sein und auch rauhes Handling oder suboptimale Umweltbedingungen ohne Datenverlust überstehen.

- b. *Lebensdauer*

Die durchschnittliche Lebensdauer des Mediums sollte bekannt sein und auch einigermaßen konstant sein. Bei den meisten modernen Medien dürfte die physische Lebensdauer des Mediums höher sein als die Lebensdauer des Systems an sich.

- c. *Prüfbarkeit*

Bei einigen Medien kann der Zustand des Mediums vor dem Beschreiben geprüft werden, z.B. durch einen Testlauf, bei dem ein bestimmtes Bitmuster auf das Medium geschrieben wird, welches danach überprüft wird (mehrmals beschreibbare Medien). Bei anderen Medien kann zwar nicht das individuelle Medium, aber das System als Ganzes geprüft werden (CD-R<sup>12</sup>).

- d. *Zuverlässigkeit, Vorhersagbarkeit des Verfalls*

Wenn die Lebensdauer eines Mediums stark variiert und deshalb schlecht abgeschätzt werden kann, ergeben sich im Bereich Langzeitarchivierung Probleme.

- e. *Verbreitung*

Wie bei anderen Kriterien ist eine hohe Verbreitung des Mediums ein grosser Vorteil, da damit eine gute und lange Unterstützung des Mediums durch die Industrie garantiert ist.

- f. *Kapazität*

Eine hohe Speicherkapazität ist von Vorteil, da damit die Anzahl der zu verwaltenden Datenträger minimal und damit auch die Menge der Ordnungsdaten klein gehalten wird. So wird z.B. bei einem Archiv von 1 TB das Handling von 10 LTO-Bändern einfacher und deshalb auch sicherer sein als der Umgang mit 1500 CD-R's.

**Frage: Welches Medium wird verwendet ?**

<sup>12</sup> Für CD-R gibt es Testsysteme, mit welchen die Kombination CD-Brenner, CD-R-Rohling und Schreibgeschwindigkeit getestet werden kann. Die Erfahrung, insbesondere aus dem Audiobereich, hat gezeigt, dass die richtige Kombination dieser drei Faktoren essentiell für eine gute Langzeitstabilität ist. Zudem sollte jede neue Charge von Rohlingen systematisch auf die Qualität kontrolliert werden.

### 3. Metadaten

#### 3.1. Verschiedene Formen von Metadaten

Metadaten sind Daten über Daten. Art und Umfang dieser Daten werden oft recht beliebig definiert, was Anlass zu Verwirrung geben kann. In Tabelle 1 sind die drei wichtigsten Kategorien von Metadaten aufgelistet<sup>13</sup>.

|   |   |
|---|---|
| <i>Beschreibende Metadaten</i>                | <i>Beschreibung des Objektes, z.B. des Bildinhaltes einer Fotografie, Ort und Zeit etc.</i>   |
| <i>Ordnungsdaten/Administrative Metadaten</i> | <ul style="list-style-type: none"> <li>• <i>Information, wo und wie das Objekt abgelegt ist.</i></li> <li>• <i>Information, wie die digitalen Daten zu interpretieren sind.</i></li> <li>• <i>Informationen zur Digitalisierung (z.B. Scanner, Auflösung etc.)</i></li> </ul> |
| <i>„Structural metadata“</i>                  | <i>Daten über den Zusammenhang bei komplexen, mehrteiligen Objekten</i>   |

#### 3.2. Grundsätze zur Archivierung von Metadaten

Ein Minimalsatz von Metadaten sollte bei der Digitalisierung erfasst und zusammen mit dem digitalen Bild abgespeichert werden. Als absolutes Minimum müssen die digitalen Daten *eindeutig* identifizierbar sein (z.B. mit einem eindeutigen, sinnvollen Dateinamen). Optimal ist es, wenn ein kleiner Satz von Metadaten direkt mit dem Bild zusammen (z.B. im TIFF-Header) gespeichert werden kann. Eine eindeutige Identifikation der Datensätze ist auch für die Qualitätskontrolle unabdingbar. Eine geeignete Möglichkeit besteht aus dem von der Newspaper Association of America (NAA) und der International Press Telecommunications Council (IPTC) entwickelten Informationsstandard zum Identifizieren übermittelter Texte und Bilder. Dieser Standard umfasst Einträge für Objektbeschreibungen, Stichwörter, Kategorien, Bildrechte und Herkunft. Bildunterschriften und Stichwörter können beim Durchsuchen von Bilddatenbanken von Drittanbietern verwendet werden (z.B. erlaubt *Adobe Photoshop®* die Eingabe nach diesem Standard).

#### 3.3. Verschiedene Standards

Im Verlauf der letzten Jahre wurden verschiedene Standards für die Erschliessung von Textdokumenten erarbeitet, die später dann auch für andere Medien ausgebaut wurden. Mit der Entwicklung des MARC (Machine Readable Catalogue) Datenformats wurde ein Gerüst geschaffen, mit dem sämtliche Metadaten

<sup>13</sup> Der "Bericht „Neue Technologien und Kulturgüter“ (R. Gschwind, L. Rosenthaler und F. Frey, Konzept: Neue Technologien und Kulturgüter, Mai 2000, erhältlich beim Bundesamt für Zivilschutz, Sektion Kulturgüterschutz, Monbijoustrasse 91, 3003 Bern) gibt einen Überblick über die verschiedenen Typen von Metadaten.

elektronisch gespeichert, gelesen und ausgetauscht werden konnten. MARC ist inzwischen ein Standard, der in den meisten Bibliotheken verwendet wird.

Im Folgenden werden einige der wichtigen Standards kurz beschrieben. Zu jedem der beschriebenen Standards ist auf dem WWW eine Fülle von Informationen zu finden.

### **MARC (Machine Readable Catalogue)**

MARC ist ein Datenformat zur elektronischen Erfassung von Medien. Es wurde von der Library of Congress in Washington DC entwickelt, um die auf Katalogkarten enthaltenen Informationen elektronisch zu verarbeiten. Die Beschreibung eines Dokumentes erfolgt in numerisch definierten Feldern. Die Information wird dadurch eindeutig identifizierbar und austauschbar<sup>14</sup>.

### **EAD (Encoded Archival Description)**

EAD nahm seine Anfänge 1993 in der Bibliothek der University of California in Berkeley. EAD wurde für die Beschreibung von Archivmaterialien entwickelt. Das Prinzip von MARC wird auf differenzierten Hierarchiestufen und komplexen Verknüpfungen von Archivalien angewendet. Die Daten werden in SGML erfasst<sup>15</sup>.

### **SGML (Standard Generalized Markup Language)**

SGML wurde 1986 als ISO Norm 8879 zur Strukturierung von elektronischen Informationen definiert. Die Daten werden so erfasst, dass ihr Inhalt und ihre Struktur eindeutig identifizierbar und plattformunabhängig sind. Für jedes Dokument wird vorab ein Typus definiert (DTD = Document Type Definition), worauf die typengerechte Strukturierung folgt. Bei Webseiten ist diese DTD HTML.

### **XML Extensible Markup Language**

1997 wurde XML entwickelt – neben EAD und HTML eine weitere Anwendung von SGML. XML bietet umfassendere Möglichkeiten als HTML, welches von XML zunehmend in den Hintergrund gedrängt wird. XML entwickelt sich immer mehr zur „lingua franca“ des Datenaustausches.

Diese Vielfalt an Methoden und Standards macht es notwendig, dass genau festgelegt wird, in welcher Form und nach welchen Richtlinien erschlossen werden soll. Erfreulicherweise hat es sich in letzter Zeit gezeigt, dass Institutionen beginnen, sich abzusprechen, und es damit möglich wird, dass Bilddaten in absehbarer Zeit auf einfache Weise nicht nur innerhalb einer Institution, sondern auch zwischen Institutionen gesucht und ausgetauscht werden können. Ein gutes Beispiel dafür bietet das europäische European Visual Archive (EVA) Projekt (<http://www.eva-eu.org>).

## **3.4. Beschreibende Metadaten**

Die Kernaufgabe der Erschliessung (mit *descriptive metadata*) besteht darin, die Bilder in einen möglichst guten Kontext zu bringen. Im Vergleich zu konventionellen Katalogen, die kein interaktives Recherchieren und Präsentieren von Bilddokumenten ermöglichen, liegt gerade darin einer der grössten Vorteile des

<sup>14</sup> Mehr Informationen können unter <http://www.loc.gov/marc/> gefunden werden.

<sup>15</sup> Die offizielle Webseite befindet sich unter <http://www.loc.gov/ead/>

elektronischen Bestands (s. Referenz 1). Der Arbeitsaufwand für die Bereitstellung und manuelle Eingabe der Daten ist nicht zu unterschätzen. Falls kein oder nur ein unvollständiger analoger Katalog vorhanden ist, benötigt dieser Teil des Digitalisierungsprojektes die grössten Ressourcen in Form von Arbeitsstunden und Geld. Es hat sich in vielen Projekten gezeigt, dass die Erschliessung abgeschlossen sein muss, bevor die Digitalisierung beginnt. Qualitätskontrolle ist äusserst wichtig in diesem Bereich, da ein Tippfehler es unmöglich machen kann, ein Bild später wieder aufzufinden.

Die Erschliessung von Bildern ist weit weniger fortgeschritten als die Erschliessung von Textdokumenten. Das hat verschiedene Gründe. Zum einen waren Bildersammlungen über lange Zeit ein Sondergebiet, das wenig Beachtung fand, zum anderen wirft die Bilderschliessung Fragen auf, deren Beantwortung zeitraubend ist und eine Fachperson erfordert.(1) Hinweis?

### **3.4.1. Ordnungsdaten/Administrative Metadaten**

Eine weitere Kategorie der Metadaten sind die Ordnungsdaten oder administrative Metadaten. Diese dokumentieren sämtliche Prozesse, die im Zusammenhang mit den Originalen und den digitalen Daten durchgeführt werden. Die folgende Liste stellt einige Beispiele administrativer Metadaten vor.

#### **Copyright**

Die Rechtslage und eventuelle Nutzungsbedingungen müssen abgeklärt und festgehalten werden.

#### **Arbeitsprotokolle**

Die Arbeitsprotokolle dokumentieren alle Abläufe. Sämtliche Arbeitsschritte vom ersten Scan bis zur Datensicherung müssen festgehalten werden, damit sie nachvollziehbar bleiben (siehe Kapitel 3.1.). Standardisierung der Arbeitsprotokolle ist die Voraussetzung, dass Daten auf verschiedenen Systemen auch in Zukunft brauchbar bleiben.

#### **Workflow**

Ein detailliertes Monitoring des Workflows ist unentbehrlich. Es ist wichtig, dass man jederzeit Informationen hat, an welcher Stelle des Workflows sich ein bestimmtes Objekt befindet. Dazu gehört auch, dass Buch geführt wird, wer wann welche Tätigkeit ausgeführt hat.

### **3.4.2. Technische Metadaten**

#### *Was sind technische Metadaten?*

Technische Metadaten sind eine Subkategorie der administrativen Metadaten. Sie beschreiben genau wie das Datenfile erstellt wurde, bieten aber auch die Möglichkeit, genaue Daten über das System, das zum Scannen benutzt wurde, abzuspeichern. Der NISO Katalog, der im Folgenden kurz vorgestellt wird, soll zeigen, welche Fülle an Daten als technische Metadaten abgespeichert werden können. Man muss sich im Klaren sein, dass nur eine gewisse Anzahl der Elemente vorgeschrieben ist. Ein grosser Teil der Elemente kann, muss aber nicht gespeichert werden.

### *NISO Metadaten-Katalog*

Vor zwei Jahren hat NISO mit dem Erarbeiten eines Metadaten Kataloges für digitale Bilder begonnen ([www.niso.org](http://www.niso.org)). Ziel war es, einen Katalog von technischen Metadaten zu erstellen, der für das Managen von digitalen Bilddaten-files notwendig ist. Der Niso Metadaten Katalog („Technical Metadata for Digital Still Images“) ist nicht nur für kulturelle Institutionen gedacht, sondern richtet sich an alle Personen, die professionell digitale Bilddaten-files von Objekten in Sammlungen erstellen.

Der Datenkatalog ist nicht abhängig von einem bestimmten Fileformat. Der NISO Metadaten Katalog kooperiert mit anderen gängigen Standards, was die Anwendung in verschiedensten Gebieten erleichtern wird. Die erste Phase des Datenkatalogs ist abgeschlossen; in einer nächsten Phase wird der Datenkatalog in einer Beta-Version von verschiedenen Institutionen ausgetestet. Diese Phase sollte spätestens Mitte 2003 abgeschlossen sein. Die Beschreibung der Elemente erlaubt es, auf einfache Weise einen DTD Katalog zu erstellen. Die Kodierung der NISO Metadaten wird in den meisten Fällen in XML erfolgen. Im Folgenden werden einige der Datenkategorien beschrieben:

- *Basic Image Parameters*  
Diese Elemente erlauben es, dass das digitale File auf einem Monitor abgebildet werden kann.
- *Image Creation*  
In dieser Sektion befinden sich alle Angaben zum Digitalisieren. Dabei ist es auch wichtig, den Namen der Person, die den Scanner bedient hat, zu dokumentieren. Diese Sektion enthält viele Daten, die automatisch beim Scannen vom Scanner aufgezeichnet werden. Unter anderem wird auch die genaue Zeit, wann der Scan gemacht wurde, aufgezeichnet. Diese Angabe sollte nach der Aufnahme nie mehr geändert werden.
- *Image Performance Assessment*  
Die Daten in diesem Abschnitt sollen die Qualität der digitalen Files garantieren. Zum einen sollen Testvorlagen dazu dienen, die Qualität der Scans zum Zeitpunkt der Aufnahme zu überwachen, zum anderen sollen sie ermöglichen, die Scans in Zukunft auf ein anderes System zu übertragen ohne Qualität einzubüßen. Damit soll insbesondere Migration als Archivierungsstrategie unterstützt werden. Die Testvorlagen entsprechen den Vorlagen, die im Teil „Qualitätskontrolle“ beschrieben wurden (siehe Kapitel ?).
- *Change History*  
Dieser Abschnitt des Datenkatalogs dient dazu, alle Prozesse (Editieren oder Transformieren), welche Bilddaten durchlaufen, zu dokumentieren. Dieser Datenblock dient nicht dazu, Prozesse rückgängig zu machen sondern die Qualität der Daten beurteilen zu können.

### *Erstellung technischer Metadaten*

Technische Metadaten werden an verschiedenen Stellen des Workflows kreiert. Eine grosse Anzahl technischer Metadaten wird beim Scannen automatisch in den File Header eingefügt (siehe Tabelle ?). Andere Daten werden bei der Nachkontrolle automatisch oder manuell zugefügt. Es ist sehr wichtig, dass diese Metadaten kontrolliert werden, um sicherzustellen, dass die Metadaten und der Inhalt der Datei übereinstimmen.

### *Minimale Metadaten*

Ein minimales Set von Metadaten sollte bei jedem Scan angefertigt werden. Man muss sich im Klaren sein, dass dieses minimale Set von Metadaten nicht als Preservation Metadata genügen kann.

### *Dublin Core*

Der Dublin Core wurde 1995 von OCLC (Online Computer Library Center) in Dublin, Ohio ins Leben gerufen. Ursprünglich wurde der Dublin Core zur Strukturierung von elektronischen Daten im Internet entwickelt. Er ist jedoch auch auf physische Daten anwendbar und eignet sich für alle visuellen Materialien. In seiner heutigen Form besteht der Dublin Core aus 15 Elementen die für die Erfassung der essentiellen Information genügen, er ist daher nicht mit den komplexen klassischen Regelwerken zu vergleichen. Dublin Core ist seit einigen Wochen ein internationaler Standard. Ausführliche Informationen sind zu finden unter [www.dublincore.org](http://www.dublincore.org).

### Die 15 Elemente des Dublin Core (Stand Februar 2002)

|             |   |
|-------------|---|
| Title       | Titel des Dokumentes  |
| Creator     | Name des Urhebers   |
| Subject     | Standardisierte Wiedergabe der Thematik                               |
| Description | Inhaltsbeschreibung   |
| Publisher   | Name des Herausgebers   |
| Contributor | Name der Person, die an der Herstellung beteiligt war                 |
| Date        | Datum, das mit dem Dokument im Zusammenhang steht                     |
| Type        | Zuordnung zu einem Genre  |
| Format      | Dokumentenspezifische Charakteristika                                 |
| Identifier  | Eindeutige Identifikation des Dokumentes                              |
| Source      | Quelle, die als Vorlage für das Dokument gedient hat                  |
| Language    | Sprache des intellektuellen Inhalts                                   |
| Relation    | Verknüpfung zu einem verwandten Dokument                              |
| Coverage    | Zeitliche oder geografische Spanne des Dokuments bzw. seines Inhaltes |
| Rights      | Informationen zur Rechtslage des Dokumentes                           |

Jedes der 15 Elemente wird seinerseits durch zehn Attribute charakterisiert. Zu den Attributen zählt beispielsweise das Standardschema, nach welchem die Information in das Element eingetragen wird. Im Weiteren wird in den Attributen festgehalten, ob ein Element obligatorisch ist, ob es wiederholbar ist und wer diese Bestimmungen festgelegt hat.

Man sollte sich auf jeden Fall vor Augen halten, dass es sich beim Dublin Core wirklich nur um ein minimales Set von Metadaten handelt. In einigen Projekten haben sich Institutionen dazu entschlossen, dieses Minimalset in einheitlicher Form zu erstellen und damit den Austausch und das sammlungsübergreifende Suchen zu ermöglichen. Oft sind aber intern sehr viel mehr Metadaten vorhanden

aus denen das Dublin Core Set „gemappt“ wird. Das EVA Projekt bietet wiederum ein sehr gutes Beispiel für dieses Vorgehen.

### *Kodierungsschema*

Es gibt verschiedene Möglichkeiten, wo und wie die Metadaten kodiert werden können. Drei verschiedene Möglichkeiten werden heute in den meisten Fällen angewendet:

- 1) Die Metadaten werden im Header der Datenfiles eingelagert. Dabei hängt es vom verwendeten Filmformat ab, wie viele Daten eingefügt werden können. Es ist klar, dass genau definiert werden muss, welche Daten sich in welchem Feld des Headers befinden. Diese Methode ist nicht geeignet für ein schnelles Abfragen der Metadaten, da das File geöffnet werden muss.
- 2) Die Metadaten werden in einer separaten Datenbank abgespeichert. Diese Datenbank enthält die automatisch im Fileheader abgespeicherten Metadaten, aber auch manuell eingegebene Daten. Es ist sehr wichtig, dass die Relation zwischen dem Datenfile und dem Eintrag in der Datenbank sichergestellt ist, z.B. dass Änderungen im Datenfile konsistent festgehalten werden.
- 3) Die Metadaten werden „on the fly“ z.B. aus einem MARC Record abgeleitet.

Es muss auf jeden Fall sichergestellt werden, dass sich die Metadaten immer auf dem aktuellsten Stand befinden, und dass die Daten im Fileheader und in der Datenbank übereinstimmen, wenn Änderungen vorgenommen werden.

Unabhängig von der gewählten Methode muss entschieden werden, wie die Metadaten kodiert werden sollen. Mehr und mehr Institutionen gehen dazu über, die Daten in XML zu kodieren. Es wird sich zeigen, wie sich das in der Zukunft entwickelt. Auf jeden Fall ist klar, dass nur ein standardisiertes Kodierungsschema den Austausch von Daten auf einfache Weise möglich machen wird.

### *Wo werden Metadaten im Arbeitsablauf eingefügt?*

Metadaten werden an den verschiedensten Stellen des Workflows eingefügt. Ein ganzes Set von *descriptiven* Metadaten sollte erstellt werden, bevor die Digitalisierung beginnt. Wird dies nicht gemacht, läuft man Gefahr, dass die Digitalisierung schneller vorankommt als die Erstellung der deskriptiven Metadaten. Dadurch kann sehr schnell ein „backlog“ entstehen.

Ein weiterer Teil der Metadaten wird während des Scanvorgangs automatisch hinzugefügt. Andere Metadaten müssen an verschiedenen Stellen des Workflows von Hand eingegeben werden. Es ist wichtig, dass die Eingabe und der Scanvorgang zeitlich nahe beieinander liegen, nur so ist zu gewährleisten, dass allfällige Fehler sofort korrigiert werden können. Es ist auch sehr wichtig, dass eine rigorose Qualitätskontrolle der Metadaten durchgeführt wird. Nur so kann sichergestellt werden, dass alle Informationen, die zur Langzeitarchivierung nötig sind, in der richtigen Form bereitstehen. Der folgende Abschnitt soll nun abschliessend noch erklären, welche Metadaten zur Langzeitarchivierung benötigt werden.

### *Metadaten zur Langzeitarchivierung*

*Preservation Metadata*, Metadaten für die Langzeitarchivierung bestehen zum grössten Teil aus administrativen und strukturellen Metadaten. Die folgenden Punkte sollten dabei beachtet werden:

- Technische Daten, die Entscheidungen für die Langzeitarchivierung unterstützen, müssen gespeichert werden
- Prozesse zur Langzeitarchivierung, wie Migration oder Emulation, müssen genau dokumentiert werden
- Die Auswirkungen der gewählten Prozesse zur Langzeitarchivierung müssen festgehalten werden
- Es muss sichergestellt werden, dass die Authentizität der digitalen Daten auf Dauer gewährleistet werden kann
- Informationen zum Copyright müssen genau und auf dem neusten Stand festgehalten werden

Die Research Libraries Group (RLG) hat 1998 das folgende Set von Metadaten für die Langzeitarchivierung publiziert (?):

Date  
Transcriber  
Producer  
Capture Device  
Capture Details  
Change history  
Validation key  
Encryption  
Watermark  
Resolution  
Compression  
Source  
Color  
Color management  
Color bar/grayscale bar  
Control targets

Diese Liste zeigt, dass es sich um administrative und strukturelle Metadaten handelt. Diese Liste ist nicht als Standard anzusehen, sie soll als Ausgangspunkt dienen. Es wird wichtig sein, Projekte zu verfolgen, die Langzeitarchivierung von digitalen Daten im grossen Rahmen durchführen.

Die folgenden Projekte sollen als Beispiele dienen, wie die Langzeitarchivierung erfolgen kann und welche Metadatenstrategien angewendet werden (für detaillierte Informationen zu den einzelnen Projekten sind die URL's als Referenzen angegeben).

- CEDARS (CURL Exemplars in Digital Archives Project)
- National Library of Australia
- NEDLIB (Networked European Deposit Library)
- Harvard University's Digital Repository Services (DRS)

Die ersten drei Projekte folgen dem OAIS (Open Archival Information System Reference Model) Modell ([www.cccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf](http://www.cccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf)).

The OAIS reference model is a conceptual framework for a digital archive. The model establishes terminology and concepts relevant to digital archiving, identifies the key components and processes endemic to most digital archiving activity, and proposes an

information model for digital objects and their associated metadata. The reference model does not specify an implementation, and is therefore neutral on digital object types or technological issues.

Harvard University's Digital Repository Services (DRS) benützt eine XML Struktur für die Preservation Metadaten.

Unabhängig von der gewählten Methode zeigen die vier Projekte, dass die Preservation Metadaten unabhängig sein sollten vom Typ des digitalen Objektes und von der Technologie, die für die Langzeitarchivierung verwendet wird. Daher kann eine einzige, einheitliche Vorgehensweise für Preservation Metadaten für viele der verschiedenen Aktivitäten der Langzeitarchivierung verwendet werden.

### 3.5. Referenzen

- 1) Kathryn Pfenninger, Bildarchiv Digital, Sonderheft 7 Rundbrief Fotografie, 2001.
- 2) Cedars project [www.leeds.ac.uk/cedars/MD-STR~5.pdf](http://www.leeds.ac.uk/cedars/MD-STR~5.pdf)
- 3) Consultative Committee on Space Data Systems (1999) Reference Model for an Open Archival Information System [www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf](http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf)
- 4) Harvard University Library (2000) Digital Repository Services (DRS): User Manual for Data Loading: A Guide for Producers of Digital Still Images (version 1.9)
- 5) National Library of Australia (1999) "Preservation Metadata for Digital Collections: Exposure Draft" [www.nla.gov.au/preserve/pmeta.html](http://www.nla.gov.au/preserve/pmeta.html)
- 6) Networked European Deposit Library (2000) "Metadata for Long Term Preservation" <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>

## 4. Richtlinien zur Digitalisierung von Fotografien

### 4.1. Einleitung

Digitalisierungsprojekte, die Digitalisierung fotografischer Sammlungen, werden in vielen Museen und Archiven geplant oder schon in durchgeführt. Die Gründe dafür können vielfältig sein, einer davon ist auch der Kulturguterhalt, die Langzeitsicherung des fotografischen Bildes, als Alternative zur fotografischen Mikroverfilmung. Oft ist der Auslöser für ein Digitalisierungsprojekt jedoch nicht primär die Langzeitarchivierung, sondern beispielsweise eine Verbesserung des Zugriffes auf die Sammlung, ein Web-Auftritt etc. Die Anforderungen eines Digitalisierungsprojektes, das Projektmanagement, der Workflow können komplex sein und müssen gut vorbereitet sein. Ein schematischer Workflow ist im Anhang dargestellt.

Wird bei einer Digitalisierung von fotografischem Bildmaterial auch eine digitale Langzeitarchivierung ins Auge gefasst, so gelten ganz spezielle Anforderungen und Qualitätsansprüche<sup>16</sup>. Wir gehen dabei davon aus, dass das digitale Bild als *Substitut* für die *visuellen Inhalt* des Originalbildes dienen soll, d.h. das digitale Bild ist für all jene Zwecke ein vollständiger Ersatz des Originalbildes, wo „nur“ der visuelle Inhalt des Bildes von Interesse ist. Daraus ergibt sich als natürliche Konsequenz das allgemeine Kriterium, mit welchem die Qualität der Digitalisierung beurteilt werden muss:

*Das „digitale Faksimile“ enthält mindestens soviel visuelle Information, wie mit konventionellen Mitteln reproduktionstechnisch aus dem Originalbild extrahiert werden kann.*

Allerdings ist eine qualitativ ausreichende Digitalisierung noch nicht ausreichend, um allen Ansprüchen eines Digitalisierungsprozess mit dem Ziel Langzeitarchivierung zu genügen. Ein solcher Digitalisierungsprozess muss in den drei folgenden Bereichen speziellen Qualitätsansprüchen genügen:

#### 1) *Digitalisierung*

Die eigentliche Digitalisierung (oder Scannen) stellt die Brücke zwischen der analogen Welt (der klassischen Fotografie) und der numerischen Welt der digitalen Reproduktion ( d.h. des maschinell verarbeitbaren digitalen Codes) dar. Dieser Übersetzungsprozess muss genügend „treu“ sein, um den visuellen Inhalt des Bildes im digitalen Code abzubilden. Schlussendlich wird hier die obere Grenze der Qualität des digitalen Bildes festgelegt. Es gibt, ausser einer erneuten Digitalisierung, keinen Prozess, welcher eine qualitativ ungenügende Digitalisierung nachträglich verbessern kann.

#### 2) *Metadaten*

Die Metadaten (Informationen über das Bild) müssen für jedes Bild erfasst und sinnvoll abgelegt werden. Ein Bild ohne die dazugehörigen Metadaten

---

<sup>16</sup> Man muss sich klar vor Augen halten: Digitalisieren zur Langzeitsicherung ist kein konservatorischer Akt, d.h. keine Konservierung des fotografischen Originals, sondern entspricht am ehesten der Mikroverfilmung.

(z.B. Zeitpunkt, Ort der Fotografie, Umstände, Beschreibungen etc.), ist für den Bereich Kulturguterhalt völlig wertlos.

### 3) *Qualitätskontrolle*

Die Digitalisierung muss von einer permanenten Qualitätskontrolle begleitet sein. Im Bereich des eigentlichen Digitalisierens und im Bereich der Metadaten können sowohl durch technische Unzulänglichkeiten als auch durch menschlichen Irrtum Fehler entstehen, welche die digitalen Bilder ungeeignet für die Zwecke der Langzeitarchivierung machen.

## 4.2. Digitalisierung

Bei der Digitalisierung sind diverse Parameter einzuhalten, denn es muss prinzipiell die „höchstmögliche“ Qualität angestrebt werden. Das Ziel ist die vollständige Übertragung der visuellen Information des fotografischen Originales in die digitale Domäne, d.h. die Erzeugung eines digitalen Masters oder des digitalen Faksimile.

Bei der Digitalisierung wird das „fotografische Bild“ in diskrete Einheiten zerlegt, in eine endliche Anzahl (meist quadratischer) Bildpunkte oder Pixel (räumliche Auflösung). Die „Helligkeit“ eines jeden Pixels wird bei der Digitalisierung ebenfalls in diskrete Einheiten zerlegt (fotometrische Auflösung). Bei der Digitalisierung entsteht nun gegenüber dem analogen Bild ein zweifacher Informationsverlust:

- Die beim Digitalisieren eingesetzten Scanner benutzen ein optisches System, d.h. das (noch analoge) Abbild der fotografischen Vorlage weist zwangsläufig einen Informationsverlust auf (Abbildungsfehler, Licht-Beugung, Farbfehler, Streulicht)<sup>17</sup>
- Die im analogen Bild vorhandene Information erfährt durch die Rasterung und die anschliessende Diskretisierung einen weiteren Informationsverlust

Es ist jedoch zu beachten, dass auch die visuelle Information auf dem fotografischen Original endlich ist. Kamera und fotografischer Film, Vergrößerungsgerät und Fotopapier haben eine begrenzte Abbildungsqualität resp. ein begrenztes Auflösungsvermögen. Deshalb ist eine endliche Qualität der Digitalisierung, welche sich der Qualität der fotografischen Vorlage anpasst, genügend. Eine höhere Qualität der Digitalisierung bringt (ausser unnötig grosser Datenmengen) keinen Vorteil und ist deshalb zu vermeiden.

Die im Folgenden ausgeführten Kriterien gelten für den eigentlichen Digitalisierungsvorgang:

### 4.2.1. Auswählen der Hardware

Bei der Auswahl der Digitalisierungshardware sind folgende 2 Punkte zu beachten:

- Analoge/elektronische Abbildungsqualität des Scanners  
Die Abbildungsqualität muss im Analogen (Abbildungsoptik, Mechanik) und im Digitalen (Elektronik, CCD, AD-Wandlung) gleichwertig sein. Mehr hierzu ist unter Punkt „Qualitätskontrolle“ zu finden.

---

<sup>17</sup> Dieser Informationsverlust tritt auch bei der Mikroverfilmung auf.

- **Physische Abstimmung auf Vorlagen**  
Das Gerät muss auf die (oft heiklen) fotografische Vorlage abgestimmt sein, so dass bei der Handhabung keine physische Beschädigung des Originals auftritt. Diese Kriterien müssen jeweils individuell abgeklärt werden. So sind z.B. bei Bild-Alben Flachbettscanner meist nicht geeignet, gewisse Materialien darf man nicht auf Glas pressen, die Erwärmung durch die Lichteinwirkung darf nicht zu hoch sein<sup>18</sup>, Glasnegative dürfen nicht auf Glas gelegt werden, ansonsten drohen sie, auf dem Glas zu „kleben“ und bei der Entfernung vom Scanner zu zerbrechen.

#### 4.2.2. Räumliche Auflösung

Die räumliche Auflösung des digitalen Bildes, d.h. die Anzahl Pixel (Breite mal Höhe), richtet sich nach der fotografischen Vorlage, d.h. nach dem „Informationsgehalt“. Dieses setzt sich zusammen aus dem Auflösungsvermögen der fotografischen Materialien und der Abbildungsleistung des optischen Systems. In der Fotografie ist das messtechnische Kriterium dafür die „Modulations-Transfer-Funktion“ (MTF). In der Praxis sind exakte Zahlen dazu schwierig zu erhalten. Bei modernen Materialien und Kameras sind Messungen vorhanden, bei historischem Material (das eindeutig schlechtere Abbildungsqualitäten aufweist) ist man auf Erfahrungswerte angewiesen.

Die Auflösung beim Scannen wird in dpi (dots per inch angegeben) und folgende minimale Auflösungen müssen eingehalten werden

|                                    |                   |
|------------------------------------|-------------------|
| Dias und Negative 35mm – 6x6       | ≥ 2700 dpi        |
| 6x9 – 4/5“                         | ≥ 2000 dpi        |
| 13x18 – 8/10“                      | ≥ 1200 - 1500 dpi |
| historische Glasnegative bis 13x18 | ≥ 1200 - 1500 dpi |
| historische Glasnegative 18/24     | ≥ 900 dpi         |
| Papierabzüge sw und color          | ≥ 500 dpi         |

#### 4.2.3. Tonwertumfang

Die „Töne“ einer Fotografie (Helligkeitswiedergabe) müssen vollständig und objektiv digitalisiert werden. In der digitalen Welt ist dabei die Anzahl der Grauwert- oder Farbstufen massgebend, welche zur Repräsentation des Tonwertumfangs zur Verfügung stehen. Die Masseinheit ist dabei Bits (pro Messwert). Mit 8 Bit können  $2^8 = 256$ , mit 12 Bit  $2^{12} = 4096$ , und mit 16 Bit  $2^{16} = 65536$  Tonwertstufen abgebildet werden. Die Bittiefe wird dabei wiederum von der Art der Vorlage, d.h. von deren Kontrastumfang bestimmt. Der Begriff „Tonwert, Grauwert oder Farbstufe“ muss natürlich definiert sein: Handelt es sich um physikalische Messgrößen wie Transmission/Reflexion (lineare Einheiten) oder optische Dichten (logarithmische Einheiten), ist eine visuelle Helligkeitsempfindung

<sup>18</sup> Der photochemische Einfluss des Lichtes selber spielt kaum eine Rolle, jedoch kann sich Vorlage oder Gerät bei langem Einsatz aufheizen und dieser thermische Effekt muss beachtet werden.

gemeint, oder handelt es sich um (relativ unnormierte) „Gamma-Korrektur“-Werte zwecks Linearisierung von Monitor-Kennlinien?

Der menschliche Betrachter ist in der Lage, gleichzeitig maximal 100 Helligkeitsstufen bei der globalen Betrachtung eines Bildes zu differenzieren, wobei der Helligkeitsumfang etwa einen Bereich von 1:100 beträgt. Alles was dunkler ist, wird als „Schwarz“ empfunden. Daraus würde sich ergeben, dass 8 Bit pro Wert ausreichend seien. Das menschliche Sehsystem hat nun aber die Eigenschaft, sich an die globalen Verhältnisse zu adaptieren, d.h. bei einem vorwiegend hellen Bild sind die ca. 100 Stufen nur auf den hellen Bereich verteilt, entsprechend bei vorwiegend dunklen Bildern im dunklen Bereich. Zudem kann mittels fotografischer Methoden (Herausvergrößerung von Details etc.) der Kontrastumfang lokal erhöht werden. Im Weiteren folgt die Helligkeitsempfindung des Menschen ungefähr einer logarithmischen Skala, während elektronische Lichtsensoren (CCD) einer linearen Skala folgen. Das digitale Faksimile muss alle diese Möglichkeiten berücksichtigen, weshalb im Allgemeinen 8 Bit pro Messwert zu wenig Auflösung bieten.

Der Tonwertumfang eines fotografischen Materials wird anhand der „charakteristischen“ Kurve bestimmt. Bei der Digitalisierung wichtig sind einerseits der Kontrast, d.h. das Verhältnis von hellster ( $D_{\min}$ ) zu dunkelster ( $D_{\max}$ ) Stelle, andererseits der Belichtungsumfang, d.h. das Verhältnis von hellster und dunkelster Stelle beim Motiv (oder Szene), das auf dem Fotomaterial aufgezeichnet werden kann<sup>19</sup>.

| Foto-Material                         | max. Kontrast | max. Belichtungsumfang |
|---------------------------------------|---------------|------------------------|
| Dias                                  | 1:1000        | 1:100                  |
| Farbnegativ<br>„Standard“ s/w-Negativ | 1:50          | 1:1000                 |
| Hist. Glasnegative                    | 1:10'000      | 1:1000                 |
| Vergröss. Papier                      | 1:100         | -                      |

Man sieht aus obiger Tabelle, dass die Variabilität sehr gross ist. Besonders problematisch beim Digitalisieren sind Negativ-Materialien, denn die visuell bildwichtigen hellen Teile sind im Negativ dunkel und bei zu geringer Graustufenauflösung beim Scanner werden diese Teile undifferenziert aufgezeichnet.

Es kann von folgenden minimalen Anforderungen ausgegangen werden, wobei je nach Vorlage weitergehende Abklärungen notwendig sind:

|                                   |          |
|-----------------------------------|----------|
| Transparentvorlagen positiv (Dia) | ≥ 12 Bit |
| Transparentvorlage negativ        | ≥ 14 Bit |
| Aufsichtvorlage (Print), linear   | ≥ 10 Bit |

<sup>19</sup> Der Belichtungsumfang, zusammen mit dem Auflösungsvermögen (Schärfe) sind ein Mass für den Informationsgehalt eines Fotomaterials

|                                   |         |
|-----------------------------------|---------|
| Aufsichtsvorlage, „logarithmisch“ | ≥ 8 Bit |
|-----------------------------------|---------|

Die Tabelle gibt dabei nur Richtwerte wieder. Dabei ist zu beachten, dass sich diese Zahlen auf effektiv gemessenen Bits beziehen. Angaben zu den effektiven Bits/Messwert müssen den technischen Angaben des Scannerherstellers entnommen werden und sollten durch regelmässige Messungen auch überprüft werden (siehe Qualitätskontrolle).

Bei der Datenspeicherung sind aus praktischen Gründen( Computerarchitektur) die Helligkeitswerte nur im 8bit oder 16bit-Format möglich (bei Color entsprechend 3x8bit = 24bit, resp. 3x16bit = 48bit. Für eine Langzeitarchivierung sind folgende Formate zu verwenden:

|                        |                      |
|------------------------|----------------------|
| Negativ                | 16 bit <sup>20</sup> |
| Dias                   | 16 bit <sup>21</sup> |
| Aufsichtsvorlage/Print | 8 bit                |

Bei Aufsichtsvorlagen ist es möglich, die z.B. intern 10-12 Bit vorliegenden Werte logarithmisch oder „Gamma“-korrigiert auf 8 Bit abzubilden.

#### 4.2.4. Kalibrierung und Farbproduktion

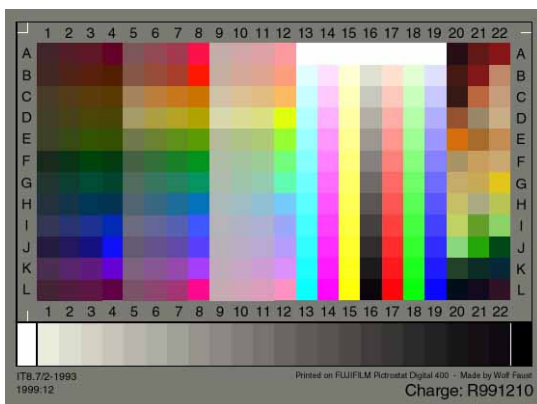
Bei der Archivierung muss man die Korrelation Pixelwert zu Helligkeit resp. Farbe kennen, d.h. der Scanner muss kalibriert werden (s. auch Qualitätskontrolle). Als Kalibriertargets stehen folgende Vorlagen zur Verfügung

Bei s/w – Vorlagen:

Graukeile (Aufsicht und Durchlicht), die bei der fotografischen Industrie bezogen werden können und die mittels eines „Densitometer“ relativ einfach ausgemessen werden können.

Bei Farbvorlagen:

a) Unter dem Standard IT 8.7 werden von mehreren Herstellern spezielle „Scanner Calibration Targets“ angeboten, sowohl für Diafilme wie für Fotopapier.



<sup>20</sup> Negative sollen als „Negative“ archiviert und nicht im Tonwert zum Positiv konvertiert werden.

<sup>21</sup> Hat man Farbdias, die keine hohen Anforderungen bezüglich Informationsgehalts stellen (z.B. Massendokumentationen), so ist eine 3x8bit Speicherung vertretbar.

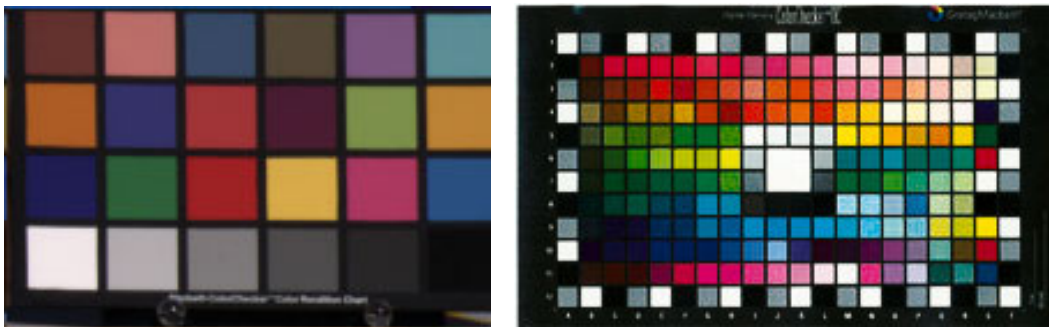
Diese Targets sind aus fotografischem Farbmaterial hergestellt und werden kalibriert ausgeliefert. Jedem Farbfeld ist der entsprechende Farbwert zugeordnet, z.B. (hier ausschnittartig dargestellt):

```

.
.
.
SERIAL "5x7 R991210"
MATERIAL "FujiFilm Pictrostat"
NUMBER_OF_FIELDS 9
BEGIN_DATA_FORMAT
SAMPLE_ID XYZ_X XYZ_Y XYZ_Z LAB_L LAB_A LAB_B LAB_C LAB_H
END_DATA_FORMAT
NUMBER_OF_SETS 288
BEGIN_DATA
A1 3.39 2.73 1.85 18.91 13.35 3.83 13.89 16.01
A2 4.58 2.98 1.63 19.97 26.04 7.92 27.22 16.92
A3 5.66 3.32 1.47 21.28 33.57 12.06 35.67 19.77
A4 5.93 3.19 1.49 20.78 38.78 10.98 40.30 15.81
A5 13.33 11.56 8.14 40.51 14.96 5.02 15.77 18.54
A6 15.27 11.21 6.98 39.93 29.46 8.65 30.70 16.37
A7 17.05 10.77 5.96 39.20 42.75 11.86 44.36 15.50
A8 21.82 10.52 4.43 38.76 68.64 18.94 71.21 15.43
A9 41.15 40.44 31.91 69.78 6.68 2.20 7.03 18.22
.
.
.

```

b) als Aufsichts-Referenzvorlage wird oft der Macbeth Color Checker resp. für digitale Kameras der Macbeth ColorChecker DC (Digital Camera)<sup>22</sup> verwendet:



Diese Targets sollen beim Digitalisieren zusätzlich mitgescannt und archiviert werden. Es ist hier wichtig, dass dies a) regelmässig geschieht und dass b) die Digitalisierung mit konstantem Scan-Parameter durchgeführt wird!

Ein ungelöstes Problem besteht bei Farbnegativ-Material, für das es kein Analogon zu den IT8 Targets gibt (und wohl auch nie geben wird). Farbnegative enthalten keine „Farben“ im visuellen Sinn, diese werden erst bei einem anschliessenden fotografischen Kopierprozess (z.B. auf Vergrösserungspapier) erzeugt, wobei es keine „objektive und eindeutige“ Farbe gibt („das Labor bestimmt die Farbe“). Die einzige Empfehlung ist, Color-Negative als 3x16bit RGB abzuspeichern. Die Umkehrung zum Positiv bleibt aber ein subjektives Unterfangen.

<sup>22</sup> Speziell für die digitale Fotografie und digitale Kameras wurde neu die Farbtafel „Macbeth ColorChecker DC“ mit 177 Farbfeldern entwickelt.

Color Management Methoden beruhen darauf, dass digitalisierte Referenz Targets (IT8.7, s.oben) mit den aktuellen Farbwerten verglichen wird. Daraus werden (durch Approximation und Interpolation) Umrechnungstabellen (Profile) erstellt. Mit diesen Profilen lassen sich die Scanner-spezifischen Eigenschaften umgehen, man erhält (mehr oder weniger) Farbinformation, die unabhängig von Scanner ist (device independent).

Für die Langzeitarchivierung sind Color Management Methoden (noch) nicht geeignet, da diese Techniken noch stark in Entwicklung sind und auch die „Standards“ laufend ändern.

Für die Kalibrierung der Grauwerte und der Farbwiedergabe wird empfohlen:

- Regelmässiges Digitalisieren von Referenz Vorlagen (Graukeile, IT8, Macbeth ColorChecker oder Macbeth ColorChecker DC)
- Digitalisieren mit konstanten Scan-Parametern

#### **4.2.5. Dunkelstrom, Rauschen (Noise) und Empfindlichkeit**

Elektronische Lichtsensoren wie z.B. CCD's unterliegen (wie jedes elektronische System) verschiedenen Störungen, deren Effekt durch eine geeignete Korrektur minimiert werden kann:

##### *1. Dunkelstrom*

Ein elektronischer Lichtsensor gibt selbst bei absoluter Dunkelheit ein kleines Signal, den sog. Dunkelstrom. Dieses Störsignal ist bei einem CCD-Array für jedes Element unterschiedlich und muss durch eine Dunkelfeldaufnahme gemessen werden. Dabei wird ein Bild ohne Beleuchtung, mit einer nicht transparenten oder schwarzen Vorlage digitalisiert.

##### *2. Rauschen (elektronisch)*

Der Dunkelstrom variiert in gewissen Grenzen zufällig. D.h., für ein bestimmtes CCD-Element wird der Dunkelstrom bei zwei aufeinander folgenden Messungen nicht identisch sein (Rauschen). Die Variabilität darf aber nicht grösser als die Soll-Genauigkeit des Scanners sein. Bei garantierten 12 Bit Auflösung darf also nur das letzte Bit zufällig variieren. Bei „nicht-optimaler“ Elektronik können zusätzlich auch periodische Strukturen auftreten, die „fixed pattern noise“ genannt werden und visuell sehr störend wirken. Im Rahmen der Qualitätskontrolle muss das Rauschverhalten periodisch geprüft werden, da sich das Rauschen durch Alterungsprozesse der Elektronik signifikant ändern kann.

##### *3. Rauschen (Quantenrauschen, Lichtrauschen)*

Bei geringer Belichtung werden pro lichtempfindliches Element eines CCD nur wenige Lichtphotonen aufgezeichnet. Auch wenn man einen „idealen“ Lichtsensor ohne Dunkelrauschen einsetzt, tritt ein (physikalisch unausweichliches) Rauschen auf, das proportional zur Wurzel der Anzahl eintreffender Lichtquanten geht, die sog. „Photonenstatistik“. Dieser Typ Rauschen kann vor allem beim Digitalisieren von dunkeln, überbelichteten Negativen ein Problem darstellen.

#### 4. *Empfindlichkeit*

Die einzelnen Elemente eines CCD-Arrays geben bei gleicher Lichtmenge nicht unbedingt das gleiche Signal. Deshalb ist eine sog. Hellfeld-Messung notwendig, bei der das Signal eines jeden CCD-Elements bei maximaler Belichtung (z.B. ohne oder mit vollständig transparenter Vorlage) gemessen wird. Damit können auch örtliche Effekte der Beleuchtung berücksichtigt werden.

Die Korrektur der Rohdaten erfolgt (für jeden Bildpunkt  $x,y$ ) dann mit der folgenden Formel:

$$I(x,y)_{cal} = \frac{(I(x,y)_{mess} - I(x,y)_{dunkel}) \cdot m}{I(x,y)_{hell} - I(x,y)_{dunkel}}$$

Wobei  $m$  den maximal möglichen Wert darstellt (z.B. bei 8 Bit Daten ist  $m=255$ ) und  $I$  den unkorrigierten Messwert.

#### 4.2.6. Geometrische Verzerrungen

Durch die Optik von Scannern können leichte geometrische Verzerrungen auftreten. Im Allgemeinen sind diese vernachlässigbar, können aber bei speziellen Vorlagen (fotogrammetrische Bilder wie z.B. Luftbilder, Bilder welche zur 3-D Erfassung von Kunstobjekten etc.) eine entsprechende geometrische Korrektur erfordern. Geometrische Verzerrungen können leicht z.B. durch Scannen von Millimeter-Papier erkannt werden.

### 4.3. Qualitätskontrolle

Bei Digitalisierungen, welche (auch) das Ziel der Langzeitarchivierung verfolgen, ist eine strikte Qualitätskontrolle auf verschiedenen Ebenen unerlässlich:

#### 1. Hardware

Die verwendete Hardware, insbesondere Scanner, ist periodisch zu überprüfen und zu eichen. Die massgebenden Parameter sollten dabei dokumentiert werden und mit vorhergehenden Messungen verglichen werden. Signifikante Änderungen der Parameter sind ein starker Hinweis auf ein Problem im Hardwarebereich (z.B. Alterung der Beleuchtung, Alterung des CCD-Chips, unzulängliche Stromversorgung und Spannungsschwankungen etc.):

- Die Korrektur von Dunkelstrom und Empfindlichkeit („Shading-Korrektur“) sollte – sofern der Scanner das nicht automatisch durchführt - idealerweise täglich, mindestens aber wöchentlich erfolgen.
- Die Eichung mit Eichvorlagen sollte ebenfalls wöchentlich erfolgen.
- Das Rauschverhalten sollte monatlich gemessen werden.

#### 2. Visuelle Kontrolle der digitalisierten Bilder

Alle digitalisierten Bilder müssen bei hoher Auflösung visuell kontrolliert werden. Besondere Beachtung muss den folgenden Merkmalen geschenkt werden:

- *Abbildungsschärfe*  
Durch mechanische Abnutzung, Dejustierung, Erschütterungen etc. kann die Geometrie der Optik verändert werden, wodurch systematische Unschärfe entstehen kann.

- *Staub und Dreck*  
Durch Staub, unsaubere Vorlagen etc. kann sich im optischen Weg des Scanners (z.B. auf der Glasplatte eines Flachbettscanners) Staub und Dreck festsetzen, welcher zu nicht tolerierbaren Störungen im digitalen Bild führt. Gerade alte Vorlagen hinterlassen oft auf dem Scanner Spuren, welche bei nachfolgenden Bildern als Störungen im Bild erscheinen und aufwändige Retuscharbeiten nach sich ziehen.
- *Geometrie*  
Sind die Bilder in der korrekten Geometrie digitalisiert und nicht spiegelverkehrt?
- *Scanfehler*  
Entspricht der visuelle Eindruck des Bildes den Erwartungen (Farbstich, Ausleuchtung, Vollständigkeit etc.)?

Die visuelle Kontrolle kann direkt während des Digitalisierungsprozesses erfolgen. Aus Effizienzgründen ist oft aber eine gründliche visuelle Kontrolle während der Digitalisierung schwierig. Die Kontrolle sollte aber häufig (mindestens täglich) durchgeführt werden, um systematische Fehler (wie z.B. Dreck auf der Glasscheibe des Scanners) so früh wie möglich zu entdecken. Die betroffenen Bilder müssen dabei erneut gescannt werden.

### 3. **Vollständigkeit und Integrität**

Die Digitalisierung von Bildbeständen mit dem Zweck der Langzeitarchivierung umfasst meistens eine grössere Anzahl von Bildern, welche zusammen ein Ensemble bilden. Durch Ermüdung, Unachtsamkeit oder andere Umstände kann es leicht vorkommen, dass einzelne Bilder „vergessen“ gehen und nicht gescannt werden, dass Dateinamen doppelt (oder wegen Tippfehler falsch) vergeben und damit das erste Bild überschrieben wird, usw. . Ist ein Ensemble digitalisiert worden, so muss unbedingt die Vollständigkeit und Integrität dieses Ensembles überprüft werden:

- Stimmt die Anzahl der digitalen Bilder mit der Anzahl der Vorlagen überein?
  - Entsprechen die Dateinamen den Erwartungen?
  - Sind die dem Bild zugeordneten Metadaten mit dem Bildinhalt konsistent?
- Diese Überprüfungen sind entsprechend zu dokumentieren.

Eine so durchgeführte Qualitätskontrolle garantiert eine Datenbasis, welche geeignet für die Langzeitarchivierung ist. Während einige Prozesse der Qualitätskontrolle nur durch eine menschliche Kontrolle durchgeführt werden können, sind andere Prozesse weitgehend automatisierbar. Bei grossen Bildbeständen empfiehlt sich eine möglichst weitgehende Automatisierung, während bei kleinen Bildbeständen eine durch menschliches Zutun durchgeführte Qualitätskontrolle ausreichend ist.

## 5. Anhang

### 5.1. Praktisches Beispiel einer digitalen Langzeitarchivierung

Am Institut für Medienwissenschaften, Abt. Bild und Medientechnologien (ehem. Abt. Wissenschaftliche Fotografie) der Universität Basel werden seit einigen Jahren sowohl Digitalisierungen als auch Langzeitarchivierung im Auftrag verschiedener Museen und Archive durchgeführt. Hier soll nun gezeigt werden, wie dabei vorgegangen wird. Die am Institut durchgeführte Langzeitarchivierung garantiert eine äusserst hohe Sicherheit der Daten, auch wenn aus diversen Gründen (Finanzierung, räumliche Einschränkungen etc.) nicht immer das ganz optimale Vorgehen gewählt werden kann.

#### 5.1.1. Digitalisierung und Qualitätskontrolle

Die Digitalisierungsparameter (Auflösung etc.) werden in enger Zusammenarbeit mit dem Auftraggeber ermittelt, um eine adäquate Qualität der digitalisierten Bilder zu erreichen. Die Bilder werden für die Archivierung im unkomprimierten TIFF-Format abgespeichert. Dabei werden die Möglichkeiten, welche das TIFF-Format bietet, weitgehend ausgenutzt, um einen kleinen Satz von Metadaten im TIFF-Header mitzuspeichern (z.B. Kalibrierung). Die eigentliche Archivierung erfolgt auf DLT<sup>23</sup> bzw. auf LTO-Ultrium Magnetbändern unter Verwendung des Programms `tar` (**t**ape **a**rchiver). `tar` ist vollkommen dokumentiert, open source, und es gibt Implementationen für alle gängigen Computersysteme. Die Aufbereitung, Archivierung und Migration der Bilddaten erfolgt auf Linux-Systemen. Der Ablauf ist wie folgt:

1. Die Bilder werden freigestellt und falls notwendig gerade gestellt (rotiert). Dieser Schritt wird teilweise interaktiv durchgeführt und dient zugleich einer ersten visuellen Kontrolle der digitalisierten Daten. Um diesen Vorgang effizient und sicher durchzuführen, wurden für diesen Zweck spezielle Programme entwickelt, welche auf einer Mischung von Interaktiver Bearbeitung und automatisierter Stapelverarbeitung beruhen.
2. Es werden die Kalibrierungsdaten erfasst und in den TIFF-Header geschrieben.
3. Ein weiterer Kontrollschritt überprüft, ob alle erwarteten Dateinamen auftreten und keine Doppelnamen bestehen. Dabei wird einerseits die Anzahl der erwarteten Dateien verifiziert, und falls möglich, die Liste der erwarteten Dateinamen mit den tatsächlich vorhandenen Dateinamen verglichen<sup>24</sup>.

---

<sup>23</sup> Es werden die beiden Typen DLT III/XT sowie DLT IV eingesetzt.

<sup>24</sup> Als Basis für diese Operationen (in Form von kleinen Shell-Skripten) dient die Unix/Linux-Funktion `diff`, welche die Differenz von Text- und Binärdateien eruiert. Es ist zu bemerken, dass diese oder ähnliche Funktionen auch in anderen Betriebssystemen zur Verfügung stehen. Für die Microsoft-Betriebssysteme gibt es z.B. Cygnus32, welches alle diese Funktionen und Skriptsprachen umfasst.

4. Im nächsten Schritt werden von den zu archivierenden Dateien mit zwei unterschiedlichen Verfahren Prüfsummen<sup>25</sup> berechnet und in ASCII-Listen abgelegt. Von diesen ASCII-Listen werden Papierausdrucke erstellt.
5. Im nächsten Schritt werden die Daten (inkl. der Checksummen) mit „tar“ auf den/die Datenträger kopiert. Danach wird der Datenträger wieder zurückgelesen und bitweise mit den Originaldaten verglichen<sup>26</sup>. Gleichzeitig wird ein Listing (Inhaltsverzeichnis) des Magnetbandes erstellt. Insgesamt werden drei identische Datensätze hergestellt, wobei zwei Datensätze auf DLT und einer auf LTO geschrieben wird.
6. Die Magnetbänder werden danach verteilt. Ein Datensatz wird dem „Kunden“ übergeben, ein Datensatz wird ausserhalb der Universität in einem Safe gelagert, während der letzte Datensatz in der Universität verbleibt.

### 5.1.2. Migration

Aktuell wird die Migration von DLT auf LTO-Magnetbändern durchgeführt. Die LTO-Technologie bietet gegenüber DLT-Technologie eine höhere Speicherdichte und ist allgemein eine modernere Technologie. Da im aktuellen Migrationsschritt nur der Datenträger, nicht aber das Bildformat (TIFF) geändert wird, ist der Migrationsvorgang relativ einfach, und kann mit einigen einfachen Skriptprogrammen<sup>27</sup> gesteuert werden:

1. Ein oder mehrere DLT-Bänder werden vollständig auf die Harddisk gelesen.
2. Es werden die Prüfsummen der Bilddateien berechnet und mit den früheren Prüfsummen verglichen. Falls die Prüfsummen übereinstimmen, wird angenommen, dass die Bilddateien korrekt auf die Harddisk kopiert worden sind<sup>28</sup>.
3. Danach werden die Daten auf LTO geschrieben und wiederum prüfgelesen. Dabei wird der Inhalt mehrerer DLT-Bänder auf ein LTO-Band geschrieben.

---

<sup>25</sup> Die Berechnung der Checksummen wird mit den Unix/Linux Funktionen `cksum` und `md5sum` durchgeführt

<sup>26</sup> Im Falle von DLT wird das Magnetband auf einem anderen Bandgerät gelesen, als es geschrieben wurde. Im Falle von LTO-Bändern können wir diesen Test (leider) nicht durchführen, da wir nur über ein LTO verfügen.

<sup>27</sup> Unter Linux bieten sich verschieden Scriptsprachen an: Es werden sowohl „tclsh“ als auch „Tcl/Tk“-Scripte verwendet.

<sup>28</sup> Um eine noch höhere Sicherheit zu erhalten, müssten zwei Bänder auf die Harddisk geladen werden und wiederum bitweise verglichen werden. Erst danach kann eine 100% Fehlersicherheit garantiert werden. Allerdings ergibt auch die Methode mit der Prüfsumme eine äusserst hohe Sicherheit.

## 5.2. Workflow-Schema zum Digitalisieren fotografischer Materialien

Folgendes Workflow Schema veranschaulicht beispielhaft, wie ein Digitalisierungsprojekt aufgebaut und durchgeführt werden soll. Ein systematischer Fragenkatalog als integraler Teil des Workflow Schemas soll es dem Leser ermöglichen, Entscheidungen im richtigen Moment zu fällen. Es ist wichtig, dass der Leser sieht, dass nicht nur ein Weg zum gewünschten Ziel führen kann.

- Selektion
  - a. Verantwortlichkeiten
  - b. Negativ / Print
    - i. Technische Implikationen
    - ii. Zustand des Materials
    - iii. Zusätzliche Überlegungen
- Copyright
  - a. Verantwortlichkeiten
  - b. Adressliste für Informationsquellen
- Wer sind die Benutzer?
- Wozu werden die digitalisierten Daten benützt?
- Welche Reproduktionsqualität wird benötigt?
- In welcher Form sollen die Daten zur Benutzung zur Verfügung stehen?
- Konservatorische Aspekte
  - a. Regelungen für Handling
  - b. Anregungen für Trainingsmodule
- Outsourcen oder in-house digitalisieren?
  - a. Welche Punkte müssen abgeklärt werden
- Digitalisierung
  - a. Parameter
    - i. Tonreproduktion
    - ii. Farbproduktion
    - iii. Auflösung
    - iv. Noise (Rauschen)
    - v. Auswählen der Hardware
    - vi. Fragenkatalog
  - b. Farbe
    - i. Farbmodelle
    - ii. Farbmanagement in einem digitalen Archiv
- Metadaten
  - a. Verschiedene Formen von Metadaten
- Technische Metadaten
  - a. Was sind technische Metadaten?
  - b. Wie werden technische Metadaten erstellt?
- NISO Metadaten-Katalog
- Minimale Metadaten
  - a. Dublin Core
    - i. Beschreibung der minimalen Elemente
  - b. Kodierungsschemen
- XML?
- Wo werden die Metadaten in den Arbeitsablauf eingefügt?

- Datenformate
  - a. Standards
  - b. Neue Entwicklungen
- Derivative
  - a. Verwendungszweck
  - b. Erstellung
- Qualitätskontrolle (wird an den nötigen Stellen im Workflowschema eingebaut)
  - a. Subjektive Qualitätskontrolle
  - b. Objektive Qualitätskontrolle
- Einrichten der Qualitätskontrollstation
- Testvorlagen mit Auswertungssoftware
  - a. Welche Standards sind im Kommen
    - i. Auflösung
    - ii. Tonreproduktion
    - iii. Farbreproduktion
    - iv. Rauschen