

Erfassung technischer Metadaten für digitale Bilder

Autoren: Robin L. Dale – [RLG \(Robin.Dale@notes.rlg.org\)](mailto:Robin.Dale@notes.rlg.org), Günter Waibel - RLG
(Guenther.Waibel@notes.rlg.org)

Einführung

Seit mehr als zehn Jahren digitalisieren die das Kulturerbe bewahrenden Institutionen aktiv ihre Sammlungen und in dieser Zeit wurden zahllose Standards – sowohl *de facto* als auch *de jure* – geschaffen, um den erweiterten, langfristigen Zugang zu diesen entstehenden digitalen Sammlungen zu erleichtern. Und während die Entwicklung der technischen Spezifikationen für Digitalisierung der Fokus war, sind einige der einflussreichsten Entwicklungen im Bereich von Metadaten geleistet worden.

Früh führte der Bedarf nach einer Verbesserung der Retrievalmöglichkeiten zur Entstehung der „[Dublin Core Metadata Initiative](#)“ und die Standardisierung von Metadaten zur Beschreibung von Dokumenten und anderen Objekten im Internet begann.

Erst 1999 kamen Fragestellungen zu technischen Metadaten bei den Institutionen auf, die sich um die Bewahrung des Kulturerbes kümmern. Obwohl die meisten dieser Institutionen begonnen hatten, ihre Sammlungen in schnellem Tempo zu digitalisieren, erhoben nur wenige durchgängig diejenigen Metadaten, welche die Funktionalität und wesentliche Qualitätseigenschaften der Bilder unter den Bedingungen langfristiger Erhaltungsstrategien gewährleisten können. Die Institutionen schrieben das Problem einem Mangel an Wissen oder fehlender Standards für die Beschreibung technischer Metadaten digitaler Bilder zu.

Im April 1999 förderten die „[National Information Standard Organisation](#)“ (NISO), der „[Council on Library and Information Resources](#)“ (CLIR) und die [RLG](#) einen Workshop, um diejenigen technischen Informationen zu untersuchen, die benötigt werden, um digitale Reproduktionen von Abbildungen, Dokumenten und Gegenständen zu handhaben und zu verwenden. Ein Ergebnis dieses Workshops war die Entwicklung von „[NISO Z39.87 Technical Metadata for Digital Still Images](#)“, ein Datenverzeichnis, das eine Standardauswahl technischer Metadaten definiert, die es Benutzern erlauben soll, digitale Bilder unter Langzeitaspekten zu erzeugen, zu interpretieren und zu verwalten.

Technische Metadaten für Bilder und NISO Z39.87

Technische Metadaten sind die erste Verteidigungslinie gegen Zugangsverlust

Warum sind technische Metadaten so wichtig? Obgleich technische Metadaten nur eine Teilmenge der für die Archivierung und die langfristige Existenzfähigkeit notwendigen Metadaten bilden, sind sie häufig als erste Verteidigungslinie gegen Zugangsverlust bezeichnet worden. Technische Metadaten stellen sicher, dass der Informationsgehalt einer digitalen Datei wieder verfügbar gemacht

werden kann, selbst wenn die gebräuchlichen, mit ihr verbundenen Dateibetrachter untergegangen sind. Außerdem liefern sie Maßstäbe, die es Maschinen wie auch Menschen ermöglicht, die Genauigkeit einer Darstellung aus einer digitalen Datei zu beurteilen.

In ihrer Gesamtheit unterstützen technische Metadaten die Handhabung und die Erhaltung digitaler Bilder während der unterschiedlichen Stadien ihrer Lebensdauer.

Technische Metadaten sind notwendig, um zwei grundlegende Funktionen zu unterstützen: die Dokumentation von Provenienz und Geschichte eines Bildes (Produktionsmetadaten) und die Absicherung, dass die Bilddaten exakt auf ein Ausgabemedium (Bildschirm, Druck oder Film) übertragen werden. Kontinuierliches Management oder „Langzeiterhaltung“ dieser Kernfunktionen setzt die Entwicklung von Anwendungen voraus, durch die Bilddaten validiert, verarbeitet, aufgefrischt und migriert werden. Dies geschieht in Abhängigkeit von Kriterien, die als technische Metadaten verschlüsselt sind.

Der NISO Z39.87 Datenkatalog erstreckt sich über vier verschiedene Kategorien:

- **Grundlegende Bildparameter** zeichnen entscheidende Informationen auf, die zum Anzeigen eines sichtbaren Bildes erforderlich sind.
- **Metadaten zur Bilderzeugung** zeichnen entscheidende Informationen für die Beschreibung der technischen Umgebung auf, in der eine digitale Bilddatei erfasst wurde.
- **Metadaten zur Bewertung der Bildqualität** zeichnen Informationen auf, die eine Auswertung der digitalen Bildqualität oder der Exaktheit der Ausgabe erlauben.
- **Metadaten zur Verarbeitungsgeschichte** zeichnen Informationen über die Verarbeitungsprozesse auf, denen ein Bild unterzogen wurde.

NISO Z39.87 steht seit 2002 für den Gebrauch unter dem Status eines „Draft Standard for Trial Use“ (DSFTU) zur Verfügung. Die Datenelemente in dieser Version bauen stark auf die Informationen auf, die in den TIFF-Dateien abgelegt werden können und obwohl nützlich, hat sich dieser Ansatz mit dem Aufkommen von Dateiformaten wie JPEG2000 als begrenzt erwiesen.

Zur Zeit unterliegt Z39.87 bedeutenden Umarbeitungen, damit es besser und genauer den Bereich der von den Institutionen gesammelten und gehandhabten Dateiformaten dokumentieren kann. Die Datenelemente der revidierten Version bauen auf der DSFTU-Versionen auf und erweitern diese. Eingeschlossen werden die in den Dateiformaten TIFF, TIFF/EP, JPEG und JPEG2000 verfügbaren technischen Metadaten wie auch Metadatenelemente des „Digital Imaging Group¹ [DIG35 metadata element set](#)“ und der [EXIF](#) Spezifikation.

Obwohl Z39.87 selbst ohne Bezug zu einer Implementierung entworfen wurde, ist das „[NISO Metadata for Images in XML Schema \(MIX\)](#)“ - von der NISO in Auftrag gegeben und erstellt von der Library of Congress - die vorherrschende Anwendung des Datenkatalogs. Da MIX ein Schema zur Erweiterung des „[Metadata Encoding and Transmission Standard \(METS\)](#)“ darstellt, war die Implementierung und der Gebrauch des Datenverzeichnisses auf lokalem Niveau relativ einfach zu handhaben. Untersuchungen haben auch gezeigt, dass Z39.87 die Erstellung lokaler Metadatensets beeinflusst und zur Ausformung umfassender Datenverzeichnisse von Metadaten beigetragen hat, die für eine Langzeiterhaltung relevant sind.²

Die Arbeitsgruppe „[OCLC-RLG Preservation Metadata: Implementation Strategies \(PREMIS\)](#)“ ist dabei, einen vollständigen Datenkatalog der Metadaten für die Langzeitarchivierung zu erstellen, der alle Dateiformate umfasst. Da technische Metadaten abhängig vom Dateiformat sind, wird durch die revidierte Version von Z39.87 eine Ergänzung zu dem übergreifenden Vorhaben von PREMIS für Datenbestände digitaler Bilder entstehen. Trotz alledem erfordert die praktische Umsetzung der Langzeiterhaltung rationale Methoden der Datenerfassung.

OCLC-RLG-Arbeitsgruppe „PREMIS“ erarbeitet vollständigen Katalog der Metadaten zur Langzeiterhaltung

Erfassung technischer Metadaten zur Unterstützung der Langzeiterhaltung

Trotz der Bedeutung der technischen Metadaten sind gemeinschaftliche Aktivitäten für die Zusammentragung der notwendigen Metadaten nur langsam in Gang gekommen. Hierfür gibt es zwei Gründe, die zueinander in Beziehung stehen: die Unfähigkeit vieler Erfassungsgeräte, einige der gewünschten technischen Metadaten aufzuzeichnen und die umfangreichen manuellen Arbeitsabläufe, auf die sich die meisten Institutionen verlassen haben, um Metadaten zusammenzutragen und zu dokumentieren.

Ein Großteil der Informationen über Erfassungsgeräte und die Aufzeichnung von Metadaten bezieht sich auf Digitalkameras und stammt aus Produktberichten oder neuesten Untersuchungen von Kodak.³ Nach diesen Beobachtungen und Berichten ist klar, dass unabhängig vom jeweiligen Standard der volle Umfang der Metadaten nicht ausreichend zunutze gemacht wird. Im besten Falle erfassen die meisten Kameras – sowohl auf Verbraucher- als auch auf professionellem Level - einige zentrale TIFF-Elemente, sehr wenige der von der Kamera gelieferten EXIF-Daten,

und wenige zusätzliche, als GPS- oder Thumbnail-Datenfelder kategorisierte Elemente. Es ist überraschend, dass geläufige „Verbraucherkameras“ geeigneter waren, Informationen aufzuzeichnen als die von derselben Firma angebotenen „professionellen“ Kameras. Jüngste Gespräche mit Mitgliedern des „[I3A/IT10 \(Electronic Still Picture Imaging\) standards committee](#)“ zeigen, dass die zunehmende Verwendung von [JPEG2000](#) (und insbesondere des JPX Dateityps) durch einige Hersteller digitaler Kameras verspricht, viel mehr technische Metadaten erfassen zu können. Diese Aussicht mag aber nur für diejenigen Hersteller zutreffen, die beabsichtigen, JPEG2000 als optionales Dateiformat für das Gerät zu realisieren. Mehr Arbeit muss getan werden, um sicher zu stellen, dass Gerätehersteller diejenigen Metadaten und Dateiformate aufzeichnen und verfügbar machen, die von der Gemeinschaft der bewahrenden Institutionen benötigt werden.

Metadatenerfassung ist arbeitsaufwendig. Die Reduktion auf ein Minimum an Metadaten ist möglicherweise kurzsichtig.

Der tatsächliche Umfang der Metadatenerfassung und der Detaillierungsgrad, mit dem die Institutionen ihre digitalen Sammlungen beschreiben, stellen ebenfalls ein bedeutendes Problem dar. Metadatenerfassung ist arbeitsaufwendig. Institutionen erheben deshalb routinemäßig minimale Metadaten, um die Kosten der Erzeugung und Speicherung ihrer digitalen Sammlung zu reduzieren. Aber diese Kosten senkende Maßnahme ist potentiell kurzsichtig. Wird eine Institution fähig sein, ihre digitalen Dateien in Zukunft wiederzugeben? Oder noch kritischer: hat die Institution genügend Informationen, um geeignete Erhaltungsmaßnahmen durchzuführen und die Informationen nutzbar zu halten? Wissen wir wirklich, *wie viele* Metadaten wir brauchen werden, um Bilddateien für lange Zeit zu erhalten?

Die Antworten auf diese Fragen sind noch nicht bekannt. Es ist unwahrscheinlich, dass wir bald genau wissen werden, wie viele Metadaten *wirklich* benötigt werden, um zukünftig Zugang und Verwaltung von digitalen Bildern zu unterstützen. Wir wissen jedoch, dass die gegenwärtige Praxis, minimale Metadaten zu sammeln, nicht ausreichend ist. Einige Institutionen haben begonnen, Erhaltungsmaßnahmen für bestimmte Bilddateien durchzuführen und man hört von der Notwendigkeit ausreichend *geeigneter* Metadaten, um die anstehenden Aufgaben durchzuführen. Aber wie kann eine Institution ausreichend korrekte technische Metadaten erwerben, um sich für die Zukunft abzusichern *und* die rationelle Erstellung von digitalen Sammlungen zu erleichtern? (Selbst Z39.87 enthält ungefähr 125 Metadatenelemente und ungefähr 40 dieser Elemente sind zwingend notwendig). Die einzig realistische und machbare Konsequenz ist: die Erzeugung und Sammlung von Metadaten soweit wie möglich zu automatisieren.

Automatische Sammlung technischer Metadaten: Die „Automatic Exposure Initiative“

Der erste Schritt hin zur Automatisierung der Sammlung technischer Metadaten ist die Identifizierung des Metadatensets, das erzielt werden soll. Der Umfang von Z39.87 erfüllt diese Rolle als ein anerkannter, aus der Anwergemeinschaft geborener, bald internationaler Standard.⁴ Das begleitende MIX-Schema dient als praktische Implementierung für Institutionen, um die Information aufzunehmen, besonders im Kontext von METS. Problem eins gelöst. Dennoch behinderten bis vor kurzem zwei Probleme den Fortschritt bei der automatisierten Metadatenentnahme. Zum einen die Unfähigkeit von Erfassungsgeräten, den Umfang der technischen Metadaten festzuhalten, die erforderlich sind, um Langzeiterhaltung und Datenverwaltung zu unterstützen. Zum zweiten die Unfähigkeit, die in digitalen Bilddateien eingebetteten Metadaten einfach freizulegen und abzuspeichern. Um beiden verbleibenden Problemen zu begegnen, gründete die RLG die „[Automatic Exposure initiative](#)“.

Die „Automatic Exposure Initiative“ hilft Institutionen, den Erfordernissen technischer Metadaten zu begegnen, indem sie eine Vielzahl an Implementierungsstrategien verfolgt. Die Initiative führt einen Dialog mit Herstellern von High-end-Scannern und Digitalkameras über die Fragestellung, auf welche Weise Ihre Produkte technische Metadaten automatisch erfassen wie auch für den Transfer in digitale Archiv- und Ressourcenverwaltungssysteme zugänglich machen können. Ferner identifiziert sie existierende oder aufkommende Technologien zur Metadatensammlung, die in Einzelinstitutionen oder von der

„Automatic Exposure Initiative“ führt Dialog mit Geräteherstellern über das automatische Sammeln und Austauschen von Metadaten

Industrie entwickelt wurden und erkundet, auf welche Weise diese Instrumente wirksam eingesetzt werden können, um der gesamten Gemeinschaft nützlich zu sein. NISO als zuständige Organisation für NISO Z39.87 ist der Co-Sponsor der Initiative, die „[Digital Library Federation](#)“ (DLF) wie auch das „[Museum Computer Network](#)“ (MCN) sagten von Anfang an ihre Unterstützung zu.

In der ersten Phase der „Automatic Exposure Initiative“ führte RLG im Juni 2003 eine informelle Befragung durch, um Interessenvertreter, geläufige Arbeitsverfahren und die in der Anwendergemeinschaft verbreiteten digitalen Erfassungsgeräte zu ermitteln. Trotz begrenzten Umlaufs wurden gut über 100 Antworten erzielt. Zusammenfassend bestätigten die Antworten, dass die Erfassung technischer Metadaten dazu neigt, ein manuelles, zeitraubendes Verfahren zu sein. Alle antwortenden Institutionen bestätigten rückhaltlos den Nutzen der Erfassung technischer Metadaten; dennoch war nur eine Minderheit in der Lage, technische Eigenschaften von Dateien selbst auf dem einfachsten Niveau zu erfassen.

Die Antworten der Umfrage brachten nachfolgend Aktivitäten in Gang, einschließlich eines [Papiers](#), das die Probleme, Lösungen und Möglichkeiten umreißt, die innerhalb der „Automatic Exposure Initiative“ erforscht werden sollen. Ferner ermittelte die Umfrage die vorherrschend benutzten digitalen Erfassungsgeräte, die in den Gedächtnisinstitutionen verwendet werden. Dabei lieferte sie auch eine Auswahlliste derjenigen Hersteller, mit denen während des Verlaufes der Initiative zusammengearbeitet werden kann. Schließlich identifizierte das Papier eine Anzahl von Softwareprogrammen, die von oder für jene Institutionen entwickelt worden sind, um ihnen zu helfen, technische Metadaten aus Bilddateien herauszulösen und zu exportieren. Die erstellte [Liste](#) stellt eine bedeutende Leistung der Gemeinschaft dar, um das für die Langzeiterhaltung notwendige Instrumentarium zu schaffen. Wichtiger noch ist, dass die meisten Instrumente unter Nutzung des Open-Source-Modells entwickelt worden sind und für den Gebrauch durch andere Kultur bewahrende Institutionen zur Verfügung stehen.

Neuer Dialog mit Geräteherstellern

Im Dialog mit den Herstellern beabsichtigt das Projekt, gemeinsame Interessen hinsichtlich der Erfassung von Metadaten und deren Verfügbarkeit für weitere Verarbeitungsprozesse, z.B. Transfer in Langzeitarchivsysteme, zu finden. Während die Anwendergemeinschaft standardisierte Metadatenelemente für die Langzeiterhaltung mit Z39.87 festgelegt hat, wurden von der Industrie eine Zahl an Initiativen in Gang gesetzt: Lieferung von digitalen Dateien, die sowohl Bilddaten als auch Metadaten enthalten und von Dateien, die in ihrem Code wesentliche Informationen über ihre Erzeugung, ihren Inhalt, Zugangsrechte etc. tragen. In einigen Fällen schlagen diese Initiativen Metadatensets vor, die für die digitale Erhaltung relevante Elemente umfassen (wie DIG35 oder EXIF). Andere schlagen spezifische oder allgemeine Transfermechanismen für selbst-beschreibende Metadaten vor (wie die XML-Box in JP2 von JPEG 2000, die JPX-Dateitypen oder Adobes erweiterbare Metadata-Plattform [Adobe XMP](#)). Die Industrie im Ganzen und die Hersteller digitaler Erfassungsgeräte haben bereits investiert, indem sie einige dieser Technologien entwickelt und implementiert haben. Jedoch zeigt eine Durchsicht der meisten dieser Initiativen, dass keine der existierenden Spezifikationen den vollständigen Metadatensatz liefert, wie er als entscheidend für die Langzeiterhaltung in NISO Z39.87 dargelegt ist.⁵

Im Verlauf der Initiative haben viele Gerätehersteller positiv auf die Einladung reagiert, an der Initiative teilzunehmen. Unter ihnen sind Betterlight, Creo/Leaf, HP, Kirtas Technologies, Kodak, Sinar Bron sowie PhaseOne. Die Unterstützung durch Experten wie Franziska Frey (Rochester Institute for Technology) und Don Williams (Eastman Kodak) ist hilfreich, um dieses durch die Anwendergemeinschaft angeregte Bestreben mit der Industrie zu verbinden. Zukünftige Hard- und Softwareentwicklungen dieser Hersteller werden Zeugnis darüber ablegen und die das kulturelle Erbe bewahrenden Institutionen sind dringend gebeten, in Kontakt mit Vertretern von Geräteherstellern zu treten und ihre Bedürfnisse hervorzuheben.

Blick in die Zukunft – Neue Dateiformate und neue Werkzeuge

Kürzlich haben einige Hersteller von Geräten und von Bildverarbeitungs-Software Pläne angekündigt, neue „Archiv“-Dateiformate zu entwickeln, wie das neuerdings eingeführte, Rohdatenformat „[Adobe Digital Negative \(DNG\)](#)“ oder jenes, welches durch die neue „[Picture Archiving and Sharing Standard \(PASS\) Group](#)“ geschaffen werden soll. Wir hoffen, dass diese Annäherung der Interessen uns die Gelegenheit bietet, mit den neuen Initiativen zusammenzuarbeiten, so dass künftige Dateiformate den vollständigen Z39.87 Metadatenset liefern können.

Gleichzeitig kann es sich die Gemeinschaft nicht leisten, auf ein „Wundermittel“ in Form eines allumfassenden Dateiformates zu warten, da dies unwahrscheinlich ist. Stattdessen sollten sich die Institutionen mit den vorhandenen Instrumenten vertraut machen, die bei der Metadatenaufnahme und -entnahme Unterstützung liefern. Die folgende [Übersicht](#) listet die zur Zeit verfügbaren Tools auf, die zur Metadatenextraktion verwendet werden können. Zusätzlich wird die RLG zwei neue Tools als Teil der „Automatic Exposure Initiative“ freigeben. Der erste Tool, die „Automatic Exposure Scorecard“, wird die für die Erfassung technischer Metadaten verfügbaren Technologien zusammenstellen und beschreiben. Kriterienkataloge werden auf der Website der „Automatic Exposure Initiative“ in den kommenden Monaten zugänglich gemacht. Das zweite Tool in Entwicklung ist eine Implementierung zur Z39.87-Adobe [Extensible Metadata Platform \(XMP\)](#). Es wird die Möglichkeiten zum Umgang mit Metadaten bei Adobe Photoshop erweitern. Adobe Photoshop ist ein allgemein genutztes Softwarepaket, das häufig in Digitalisierungsverfahren für Objekte des Kulturguts eingesetzt wird. Sobald es fertig gestellt ist, wird dieser Tool in einer zukünftigen Ausgabe der RLG DigiNews angekündigt und frei zugänglich auf der [RLG](#) Website sein.

¹ Die „Digital Imaging Group“ (DiG) fusionierte im July 2001 mit der „Photographic and Imaging Manufacturers Association“ (PIMA), zur „International Imaging Industry Association“ (I3A)

² Sowohl die Umfragen der „Automatic Exposure Initiative“ als auch diejenige der PREMIS-Arbeitsgruppe / Untergruppe Implementierung stellten diese erweiterte Verwendung des Z39.87-Datenverzeichnisses fest.

³ [Automatic Exposure: Capturing Technical Metadata for Digital Still Images](#). Mountain View, CA: RLG, 2004. Siehe Appendix 3: *Professional Camera Metadata Survey* von Kodak (2002), Appendix 4: *Consumer Digital Camera Metadata Survey* von Kodak, 2002, und Appendix 5: *Consumer Digital Camera Metadata Survey* von Kodak, 2003.

⁴ Obwohl das Akronym „NISO“ richtig als „National Information Standards Organisation“ übersetzt wird, werden die von dieser Organisation geschaffenen Standards größtenteils in vernetzten Umgebung benutzt. Die Beschränkung auf „national“ ist in diesem Sinne unrichtig. Tatsächlich haben die von dieser Organisation geschaffenen und unterstützten Standards internationalen Rückhalt und Einfluss.

⁵ Obwohl keine der existierenden Spezifikationen den vollständigen in NISO Z39.87 umrissenen technischen Metadatenset liefert, decken alle von ihnen mindestens einige der Datenelemente ab, die in NISO Z39.87 spezifiziert sind.

Mit freundlicher Genehmigung von RLG, Research Libraries Group, Robin L. Dale und Günter Waibel.

Original erschienen in: [RLG DigiNews, Vol 8 \(Oct 2004\), No 5](#)

*Deutsche Übersetzung von Andrea Blochmann und Hans Liegmann,
nestor - Kompetenznetzwerk Langzeitarchivierung c/o Die Deutsche Bibliothek
[\[http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20041217013\]](http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20041217013)*