



Audiovisuelle Kulturgüter erhalten

Kolloquium *Datenkomprimierung bei Bild, Audio, Video*

KOST / Memoriav

11. November 2009, Universität Bern

Inhalt

1. Vorwort.....	1
2. Einführung: Eine Begriffsklärung.....	1
3. <i>Best Practices</i> im audiovisuellen Bereich.....	2
4. Kurzstatement: Archivische Gründe für den Verzicht auf Komprimierung.....	3
5. Fallstudien im Bildbereich.....	3
6. Schlussfolgerungen.....	5

1 Vorwort

In archivfachlichen Diskussionen herrscht ein weitgehender Konsens zur Datenkomprimierung: Sie wird prinzipiell kritisch gesehen, verlustbehaftete Komprimierung grundsätzlich abgelehnt. Dieser Konsens hat sich in den verschiedenen Katalogen und Empfehlungen zu archivtauglichen Dateiformaten niedergeschlagen. Auch die KOST schloss sich in ihrem Katalog archivtauglicher Dateiformate (<http://www.kost-ceco.ch/wiki/whelp/KaD/>) dieser Übereinkunft an.

In den letzten Jahren wurde diese Position jedoch im Bereich der Bildformate verschiedentlich kritisch hinterfragt. Die Fragestellung konzentrierte sich dabei hauptsächlich auf zwei Aspekte:

- Bei Digitalisierungen von Bildern, Codices, Urkunden etc. ist der Begriff "Verlust" genauer zu klären. Ein erster, unwiederbringlicher Datenverlust findet nämlich beim Digitalisierungsvorgang statt. Darauf folgende Datenkomprimierungen beziehen sich immer auf den digitalen Master, nicht auf das analoge Original. Datenkomprimierung kann also nicht losgelöst von den Parametern der Digitalisierung und von entsprechenden *best practices* gesehen werden.
- In verschiedenen Digitalisierungsvorhaben der letzten Jahre wurde vom bisherigen archivischen Grundsatzentscheid für TIFF abgewichen. Erwähnt seien die Überlegungen der Königlichen Bibliothek (KB) in den Niederlanden sowie das Projekt MONOLITH des Staatsarchivs Zürich. Es zeigt sich, dass bei sehr grossen Datenmengen der Speicherplatzbedarf unkomprimierter TIFF-Dateien zusehends als Problem wahrgenommen wird.

Um einen Überblick über diese aktuellen Diskussionen zu geben und Archivinstitutionen in die Lage zu versetzen, sich den Herausforderungen dieser Fragestellung zu stellen, veranstaltete die KOST gemeinsam mit Memoriav ein Kolloquium zur Datenkomprimierung. Der Fokus lag dabei auf Bilddaten, da die Komprimierung in diesem Bereich besonders umstritten ist. Zwei Fallstudien beleuchteten diese Thematik. Einleitend wurden die verwendete Terminologie erläutert, die (inhaltlich weitgehend unbestrittene) Situation in den Bereichen Audio und Video kurz vorgestellt und Überlegungen aus der Arbeit am archivischen Formatkatalog des BAR präsentiert¹.

2 Einführung: Eine Begriffsklärung

Martin Kaiser, KOST²

Datenkomprimierung / Datenkompression / Datenreduktion

Die Begriffe werden synonym verwendet. "Datenreduktion" beschreibt den Vorgang am besten: Der Datenstrom (*bitstream*), der eine Datei ausmacht, wird reduziert; somit wird die Grösse der Datei kleiner. Der Gewinn an Speicherplatz hängt von der Art der Komprimierung und vom verwendeten Algorithmus ab.

Verlustfrei / verlustbehaftet

Datenkomprimierung ist grundsätzlich auf zwei Arten möglich. Bei der verlustfreien Komprimierung werden im Datenstrom vorhandene Redundanzen eliminiert (bzw. die Darstellung redundanter Daten abgekürzt).

¹ Für den Bildbereich waren drei Fallstudien vorgesehen. Leider musste der Beitrag von Christophe Brandt (*Institut Suisse pour la conservation de la photographie, ISCP*), "La numérisation, une étape sensible à intégrer dans la préservation des collections photographiques", kurzfristig ausfallen. Wir danken Kurt Deggeller (Memoriav) und Urs Meyer (BAR), die sich bereit erklärten, mit einem ausgebauten Beitrag bzw. mit einem Kurz-Statement einige Elemente der Argumentation von Herrn Brandt in das Kolloquium einzubringen.

² <http://www.kost-ceco.ch/>

Dies geschieht durch Lauflängenkodierung, statistische Kodierung (*Huffman Code*) oder andere Verfahren³. Der ursprüngliche Datenstrom kann bei der Dekomprimierung wieder hergestellt werden; die Komprimierung ist umkehrbar, es geht keine Information verloren. Bei der verlustbehafteten Komprimierung werden (ausschliesslich oder zusätzlich) solche Daten entfernt, die für die menschliche Wahrnehmung wenig oder gar nicht relevant sind. So reagiert beispielsweise das menschliche Auge auf Helligkeitsunterschiede viel empfindlicher als auf Farbunterschiede; der Verzicht auf gewisse Farbinformationen ist deshalb nicht oder kaum wahrnehmbar. Solchermassen entfernte Daten sind bei der Dekodierung jedoch nicht mehr rekonstruierbar, weswegen es sich um einen Datenverlust handelt.

Visually lossless

Die Begriffe "verlustfrei" und "verlustbehaftet" beziehen sich also auf den Datenstrom, der die entsprechende Datei ausmacht. Um auszudrücken, dass ein Datenverlust im Datenstrom nicht zwingend einen Informationsverlust im vom Datenstrom repräsentierten Informationsobjekt bedeutet, wird häufig der Begriff *visually lossless* verwendet. Diese Betrachtungsweise fokussiert also nicht auf die Daten, sondern auf die (in diesem Fall optische) Information.

Digitalisate / originär digitale Daten

Die Diskussion um Daten- und Informationsverluste kann nicht unabhängig davon geführt werden, welches Objekt am Anfang stand. Zu unterscheiden ist insbesondere zwischen originär digitalen Daten und Digitalisaten. Erstere sind digital entstanden (*born digital*), und es liegt kein analoges Original vor. Letztere sind das Produkt des Digitalisierungsprozesses eines analogen Originals. Bei der Digitalisierung werden kontinuierliche physikalische Grössen von Messgeräten zunächst als elektrische Spannung ausgegeben und dann zu diskreten Zahlenwerten digitalisiert. Parameter im Prozess sind die Abtastrate bei der Erfassung der Messwerte und die Auflösung bei der Umwandlung in digitale Werte.

3 Best Practices im audiovisuellen Bereich

*Kurt Deggeller, Memoriav*⁴

Am Beginn des Referats stand eine grundsätzliche Überlegung zur Digitalisierung und Datenkomprimierung: Es ist immer davon auszugehen, dass das Digitalisat das Original dereinst wird ersetzen müssen. Deshalb muss es Ziel der Digitalisierung sein, das Original so getreu wie möglich abzubilden. Gedächtnisinstitutionen fordern in ihren berufsethischen Regeln, dass au-

diovisuelle Dokumente das Recht haben, nach denselben Regeln behandelt zu werden wie traditionelle Materialien. Problematisch ist, dass audiovisuelle Archivierung wegen der grossen Datenmengen kostenintensiv ist, und dass Markttechnologien nicht auf Langzeiterhaltung ausgerichtet sind.

Im Bereich der Audiodaten ist der Übergang zur *file*-basierten Archivierung weitgehend vollzogen. Es existiert seit einigen Jahren kein archivtauglicher Datenträger mehr. Dafür hat sich das WAVE-Format für die Archivierung von Audiodateien auf breiter Front durchgesetzt und wird in seltener Einmütigkeit als archivtaugliches Format anerkannt⁵. Die dadurch bedingten Datenmengen sind zwar beträchtlich, werden aber kaum in Frage gestellt⁶. Das *IASA Technical Committee* hält eindeutig fest: "Datenreduktion erlaubt nicht die Wiedergewinnung der originalen akustischen Verhältnisse und schränkt überdies die weitere Verwendung der Aufnahme durch das Entstehen von Artefakten beim wiederholten Kopieren ein."⁷

Im Videobereich ist die Situation bei weitem unklarer: Die Standardisierung ist weniger weit, die Formatvielfalt ist gross, aufgrund der enormen Datenmenge sind datenreduzierte Formate kaum zu vermeiden. Klar ist einzig, dass bei der Filmdigitalisierung weniger Zeitdruck herrscht: Rettungsdigitalisierungen sind noch nicht flächendeckend notwendig. Im Videobereich ist die Obsoleszenz jedoch weiter fortgeschritten. Grosse Hoffnungen werden in das noch wenig verbreitete Format Motion JPEG 2000 (MJPEG2000) gesetzt, das rein spatiale und verlustfreie Komprimierung erlaubt und in der Filmindustrie bereits breit verwendet wird. Auf jeden Fall soll die Zusammenarbeit mit professionellen Dienstleistern gesucht werden.

In der Diskussion wurde der Blick zusätzlich auf die Tatsache gelegt, dass die Qualität jeglicher Folgeversion von derjenigen der Vorgängerversion gegen oben begrenzt wird; oder plakativ ausgedrückt: Sünden bei der Aufnahme können nicht rückgängig gemacht werden. Zudem wurde der Erwartung Ausdruck verliehen, dass die befürchtete Kumulation von Artefakten durch neue Formate gebremst oder vermieden werden kann:

⁵ In der anschliessenden Diskussion wurde präzisiert, dass WAVE nur ein Containerformat ist, das verschiedene Arten von Datenströmen beinhalten kann, auch komprimierte. Als archivtauglich wird die Pulse Code Moderation PCM angesehen, weswegen präziser vom "WAVE PCM-Format" gesprochen wird.

⁶ Als Richtwert kann die Datenmenge pro Stunde Ton für verschiedene Formate und Kompressionen dienen:

WAV-File, Stereo, 96 kHz / 24 bit	2 GB
MP3-File, Stereo, 128 kBit/s, 44 kHz	50 MB
MP3-File, CD-Qualität (256 kBit/s, 44 kHz)	100 MB

⁷ IASA Technical Committee: *Standards, Praxisempfehlungen und Strategien. IASA-TC 03. Die Bewahrung von Schallaufnahmen. Ethische Aspekte, Prinzipien und Strategien* (³2005), p. 9. http://www.iasa-web.org/downloads/publications/TC03_German.pdf

³ Cf. die detaillierten Ausführungen in der Präsentation von Peter Fornaro.

⁴ <http://www.memoriav.ch/>

Für verbreitete Videoformate kann gefahrlos eine Lebensdauer von zehn bis zwanzig Jahre angenommen werden; es ist davon auszugehen, dass bei einer dann-zumal nötigen Migration deutlich performantere Formate zur Verfügung stehen werden.

4 Kurzstatement: Archivische Gründe für den Verzicht auf Komprimierung

Dr. Urs Meyer, Schweizerisches Bundesarchiv⁸

In seinem Kurzreferat als Ersatz für den Vortrag von Christophe Brandt (ISCP) begründete Urs Meyer, weshalb das Schweizerische Bundesarchiv bei seinen Archivformaten auf Komprimierung grundsätzlich verzichtet. Archivformate sollten möglichst langlebig sein, damit der Rhythmus der Migrationen möglichst tief gehalten werden kann. Ein Beitrag zur Langlebigkeit ist eine möglichst tiefe Komplexität der Formatdefinition. Da Komprimierung Komplexität addiert, wird sie als Problem gesehen. Sollte in Zukunft Datenkomprimierung dennoch unvermeidlich werden, sind an die Komprimierungsalgorithmen die gleichen Anforderungen zu stellen wie an archivtaugliche Formate, namentlich Offenlegung und Verbreitung.

In der Diskussion wurde zudem darauf hingewiesen, dass weniger Komplexität auch robustere Formate bedeutet. Bei herkömmlichen komprimierten Formaten wie JPEG reichen Fehler von einem oder wenigen Bits aus, um das Bild stark zu verändern oder vollends unlesbar zu machen, währenddem sich der Fehler bei unkomprimierten Formaten auf die fehlerhaften Stellen beschränkt. JPEG2000 wiederum benutzt einen robusteren Komprimierungsalgorithmus, womit die angesprochenen Probleme vermieden werden können.

5 Fallstudien im Bildbereich

5.1 Bild-Datenkomprimierung aus Sicht der Nutzer: Experimente und Befunde aus der Wissenschaftspraxis

*Ralf Stockmann,
Göttinger Digitalisierungszentrum GDZ,
Staats- und Universitätsbibliothek Göttingen⁹*

Ralf Stockmann berichtete aus seiner Erfahrung mit der Massendigitalisierung von historischen Buchbeständen. Die Deutsche Forschungsgemeinschaft DFG, einer der wichtigsten Geldgeber im Bereich Digitalisierungsprojekte in Deutschland, schreibt für von ihr geförderte Projekte die verlustfreie Speicherung der Digitalisate im TIFF-Format vor. Der daraus resultierende explodierende Speicherplatzbedarf ist für das GDZ ein grosses Problem¹⁰, zumal Archivspeicherplatz (im Gegensatz

zu handelsüblichen Festplatten) in den letzten zwei bis drei Jahren kaum günstiger geworden ist. Entsprechend gross ist das Interesse an komprimierten Dateiformaten, insbesondere JPEG2000. Dieses muss wegen seiner Verbreitung, der eingeschränkten Renderingmöglichkeiten sowie der ungewissen Zukunft noch mit Vorsicht betrachtet werden. In seinem Referat fokussierte Stockmann jedoch auf die Bildqualität und stellte die These auf, dass Verfälschungen durch Komprimierung kleiner sind als die durch den Digitalisierungsprozess erzeugte Varianz.

Vergleich TIFF – JPEG2000

Als Illustration dieser These dienten im GDZ erstellte Scans historischer Bücher, und zwar sowohl in der vom Scanner gelieferten TIFF-Version als auch in einer Serie von JPEG2000-Versionen mit ansteigenden Komprimierungsraten (und entsprechend linear geringerer Dateigrösse). Ein erster, optischer Vergleich zeigte, dass bei sämtlichen JPEG2000-Versionen bis zu einer Komprimierungsstufe von 40 bei einer Darstellung in Originalgrösse kein Unterschied erkennbar war. Erst bei Vergrösserung wurden Unterschiede sichtbar; bis zu einer Komprimierungsstufe von 90 waren die Bilder aber selbst bei achtfacher Vergrösserung optisch verlustfrei.

Ein quantitativer Vergleich wurde mittels Differenzmessung durchgeführt. Dabei wird bei zwei Bildern Pixel für Pixel verglichen. Die Unterschiede können als Schwarz-Weiss-Werte dargestellt oder in einem Differenzwert von 0 (keine Differenz) bis 100 (maximale Differenz) ausgedrückt werden. Bei den komprimierten Versionen steigt der Differenzwert mit zunehmender Komprimierungsstufe an, wobei ein grosser Schritt von Komprimierungsstufe 70 zu 60 festzustellen ist. Somit scheint sich ein Idealwert für die Komprimierungsstufe von 70 zu ergeben, der eine Speicherplatzersparnis um den Faktor 3,6 bei minimalem messbarem Verlust ergibt.

Zwei gleiche Scans

Springender Punkt der Demonstration war jedoch die Messung der Differenz zweier "identischer" Scans, d.h. zweier mit dem gleichen Scanner von der gleichen Buchseite unmittelbar hintereinander erzeugter Scans. Dieser Vergleich ergab einen Differenzwert von 82, mithin deutlich höher als zwischen unkomprimiertem und komprimiertem Bild. Der Vorgang des Digitalisierens ist also weder vorhersagbar noch reproduzierbar, sondern mit einer fundamentalen Unschärfe behaftet. Auffallend war auch, dass sich im Vergleichsbild Bildmotive abzeichneten, was auf ein Schärfen der Konturen schliessen lässt. Die Interpretation beginnt also bereits im Scanner bei der Umrechnung des RAW-Bildes ins TIFF-Format. (Interessanter wäre es deshalb, auf das RAW-Bild direkt zugreifen zu können, was aber technisch unmöglich ist.)

⁸ <http://www.bar.admin.ch/>

⁹ <http://www.sub.uni-goettingen.de/>

¹⁰ Allein für Neudigitalisate im Jahr 2010 sieht die Planung des GDZ 115 TB Speicherbedarf vor.

Headroom

Ein zusätzlicher Test mass den in den Scans enthaltenen *Headroom* (Aussteuerungsreserve), d.h. die im Bild enthaltenen, aber auf den ersten Blick unsichtbaren Nuancen, die mit speziellen Bildbearbeitungsverfahren sichtbar gemacht werden können. Dazu wurde eine unbedruckte Buchseite gescannt. *Headroom*-Bearbeitungen sowohl des TIFF- als auch des JPEG2000-Bildes brachten durchscheinende Motive und Wasserzeichen zum Vorschein, und zwar ohne relevanten Unterschied.

Zusammenfassung

Aus den gezeigten Beispielen kann gefolgert werden, dass die Qualitätseinbussen einer massvollen JPEG2000-Komprimierung von Digitalisaten mit Blick auf die inhärenten Varianzen im Scanprozess selber vernachlässigbar sind. Für den Entscheid zwischen TIFF und JPEG2000 ist die Bildqualität in solchen Fällen also kein Kriterium; er muss sich einzig auf die "Archivqualitäten" dieser Formate abstützen (Verbreitung, Zukunftsperspektiven). Für Massendigitalisierungen kann nach den Erfahrungen von Ralf Stockmann die Verwendung von JPEG2000 empfohlen werden.

In der Diskussion wurde zum *Headroom*-Test angemerkt, dass seine Resultate wohl vorwiegend der Scantiefe von 8 Bit geschuldet waren. Es ist zu vermuten, dass eine Scantiefe von 16 Bit grössere Unterschiede zwischen TIFF und JPEG2000 ergeben hätte. Zur Sprache kam anschliessend die OCR-Erkennung von gescanntem Text. Stockmann berichtete, dass sich diese in vielen Fällen irrational verhält. Hier besteht noch Testbedarf. Fortschritte werden vom neuen EU-Projekt IMPACT erwartet, an dem die SUB mitarbeitet¹¹.

Auf die Frage nach dem Potential von PDF zur Bildarchivierung wies Stockmann darauf hin, dass dieses nur ein Container ist, das diverse komprimierte und unkomprimierte Bildformate enthalten kann. Für die Bildarchivierung macht die Verwendung von PDF keinen Sinn. Aus dem Publikum kam schliesslich der erneute Hinweis, dass sich in der Filmarchivierung das von JPEG2000 abgeleitete Format Motion JPEG2000 (MJPEG2000) zum Standard entwickle.

5.2 Qualitätskriterien digitaler Bilder im Langzeitarchiv

Dr. Peter Fornaro

Imaging&Media Lab, Universität Basel¹²

Peter Fornaro unterstrich die Aussagen von Ralf Stockmann, indem er die Qualitätskontrolle des gesamten Scanning- bzw. Fotografierprozesses von der Aufnahme über die Bearbeitung bis zur Archivierung darstellte. Diese umfasst auch vordergründig triviale, aber entscheidende Elemente wie die regelmässige Kontrolle der Geräte und die Vergabe eindeutiger Dateinamen.

¹¹ Cf. <http://www.impact-project.eu/>.

¹² <http://www.iml.unibas.ch/>

Im Kontext der Fragestellung dieses Kolloquiums ist insbesondere der Hinweis wichtig, dass die Komprimierung bereits bei der Digitalisierung beginnt, die ja nichts anderes ist als eine Reduktion eines kontinuierlichen analogen Originals auf einen diskreten Datenstrom in einer bestimmten Auflösung des Raumes (Pixel) und der Intensität (Helligkeit, Quantisierung).

Grundlagen

Fornaro legte ein besonderes Augenmerk auf die Farbwiedergabe und -wahrnehmung. Farben werden in der digitalen Welt durch eine Mischung dreier Grundfarben definiert, ausgedrückt in drei Zahlen. Der Farbraum gibt diesen Zahlen eine physikalische Bedeutung. Durch die Werte der Grundfarben sind verschiedene Farbräume definiert, die sich durch ihren *Gamut*¹³ unterscheiden, die Menge aller möglichen Farben. Der *Gamut* kann optisch innerhalb des maximal möglichen Farbraumes durch ein Dreieck dargestellt werden, dessen Ecken durch die Grundfarben definiert sind. Zu beachten ist, dass der durch das Internet am weitesten verbreitete Farbraum, sRGB, einen relativ beschränkten *Gamut* hat und viele Farben nicht abbilden kann; es existieren aber entsprechende Erweiterungen. Besonders bei grossen Farbräumen ist eine Quantisierung durch 256 Helligkeitsstufen (8 Bit) nicht genügend; es sollte wenn möglich eine grössere Farbtiefe gewählt werden¹⁴.

Datenreduktion

Die RGB-Codierung ist durch die Kamera- bzw. Sensortechnik bedingt. Ein Sensor nimmt Helligkeitsstufen wahr; mit davorgeschalteten Filtern wird diese Wahrnehmung auf einzelne (Grund-)Farben beschränkt; die drei Farbelemente ergeben dann die Mischung. Der Mensch nimmt Farben jedoch anhand der Elemente Helligkeit, Farbton und Farbsättigung wahr. Wenn es gelingt, den Darstellungsraum an diese Wahrnehmung anzupassen, steigt der Informationsgehalt. Zudem eröffnen sich neue Möglichkeiten zur Datenreduktion: Auf Helligkeitsunterschiede reagiert das menschliche Auge deutlich feiner als auf Farbunterschiede. Deshalb können die Farbinformationen in niedrigerer Auflösung gespeichert werden. Ein Beispiel ist das YCbCr-Farbmodell, das in JPEG-Bildern verwendet wird: Die Farbabweichungssignale Cb und Cr werden in reduzierter Auflösung gespeichert. Das JPEG-Format umfasst zusätzlich eine Diskrete Cosinus-Transformation (DCT), d.h. eine Zerlegung des Bildes in Frequenzbänder. Eine Quantisierung dieser Frequenzbänder führt

¹³ *Gamut*, englisch für Tonleiter, Skala, Farbpalette.

¹⁴ In der Diskussion wurde richtigerweise darauf hingewiesen, dass eine 16-Bit-Quantisierung selbstredend eine Verdoppelung der Datenmenge bewirkt. Je nach Ausgangsmaterialien ist dies ungerechtfertigt; allenfalls kann eine Quantisierung von 16 Bit während der Bearbeitung mit späterer Reduktion auf 8 Bit in Betracht gezogen werden (mit dem entsprechenden Ressourcenaufwand bei der Bearbeitung, z.B. für die Bearbeitungszeit).

zur Datenreduktion; dabei entstehen die typischen Kompressionsartefakte.

Das von einem digitalen Sensor erzeugte RAW-Bild ist proprietär und muss zunächst bearbeitet werden, bevor es archiviert oder sonst wie weiterverwendet werden kann. In RAW-Bildern sind die vom Sensor wahrgenommenen Helligkeitswerte für jede der drei Grundfarben gespeichert; die Farbinformation muss also zunächst in einem relativ aufwendigen Prozess errechnet werden. Deswegen und wegen ihres proprietären Charakters eignen sich RAW-Formate nicht für die Archivierung. Als Archivierungsformate in Frage kommen TIFF – ein flexibles Containerformat, sofern die empfohlenen Parameter gewählt werden – und JPEG2000, das den Spagat anstrebt zwischen Komprimierung und Archivtauglichkeit.

6 Schlussfolgerungen

Im Kolloquium wurden die aufgeworfenen Fragen rund um die Datenkomprimierung ausführlich beleuchtet, im Detail analysiert und kontrovers diskutiert. Die folgenden Punkte scheinen der KOST erwähnenswert als Erkenntnisse aus den Präsentationen und der Diskussion:

- Die Formatauswahl und Datenkomprimierung dürfen nicht als isolierte Probleme betrachtet werden, sondern müssen im grösseren Zusammenhang der Qualität bei der digitalen Archivierung und im Digitalisierungsprozess gesehen und analysiert werden.
- Am Anfang eines Digitalisierungsvorhabens muss die Frage nach dem Sinn und den damit verfolgten Zielen stehen. Eine Antwort darauf kann auch der (momentane) Verzicht auf die Digitalisierung sein, z.B. wenn aus konservatorischen Gründen noch kein Handlungsdruck besteht. Für gewisse analoge Formate wiederum drängt sich eine Rettungsdigitalisierung auf.
- Oberstes Ziel muss sein, qualitativ hochstehende und möglichst langzeittaugliche Digitalisate zu erzeugen sowie diese über lange Frist zu erhalten. Die Qualität des Ausgangsmaterials spielt dabei eine wichtige Rolle, kann jedoch nicht beeinflusst werden. Beeinflusst werden können aber sämtliche Schritte des Digitalisierungsprozesses.
- Die Digitalisierung ist immanent eine Datenreduktion. Zudem handelt es sich nicht um einen exakten, reproduzierbaren Prozess, sondern um einen mit inhärenten Unschärfen behafteten.
- Die Frage nach dem geeigneten Dateiformat für Bilddigitalisate – konkret die Frage nach TIFF oder JPEG2000 – muss anhand der Anforderungen an archivtaugliche Dateiformate beantwortet werden. Es wurde gezeigt, dass JPEG2000 mit massvoller verlustbehafteter Komprimierung aus qualitativen Gründen zu keinen Sorgen Anlass gibt. Allerdings sprechen die Stabilität und die Verbreitung mindestens im Augenblick noch für TIFF. Als Leitlinie kann gelten: Für Massendigitalisierungen kann JPEG2000 als tauglich gelten; bei Rettungsdigitalisierungen und der Erstellung von digitalen Masterversionen ist die Verwendung von TIFF in Betracht zu ziehen.
- Für originär digitale Bilddaten gelten sinngemäss die gleichen Überlegungen. Technisch besteht kein Unterschied zwischen dem Sensor einer Digitalkamera und demjenigen eines Scanners. Referenzgrösse ist dabei das digitale Original, strenggenommen das RAW-Bild der digitalen Kamera.
- Im Bereich der Audiodaten ist das unkomprimierte Format WAVE PCM praktisch unbestrittener Archivstandard. Im Bereich der Videodaten scheint es im Moment kaum möglich, auf Datenkomprimierung zu verzichten. Ein allgemein anerkanntes Archivformat besteht noch nicht. Gute Prognosen werden jedoch für MJPEG2000 gestellt.