

Dateiformate

Eigenschaften und Eignung für die Archivierung elektronischer Unterlagen

Eine Handreichung für Archivarinnen und Archivare

Auf der Grundlage der Internet-Präsentation eines Workshops des Vereins
Schweizerischer Archivarinnen und Archivare vom 16.5.2000

herausgegeben, übersetzt, erweitert und bearbeitet

von Bernward Helfer und Karl-Ernst Lupprian

Wiesbaden und München

2004

Dateiformate. Eigenschaften und Eignung für die Archivierung elektronischer Unterlagen.
Eine Handreichung für Archivarinnen und Archivare.

Auf der Grundlage der Internet-Präsentation eines Workshops des Vereins Schweizerischer Archivarinnen und Archivare vom 16.5.2000 herausgegeben, übersetzt, erweitert und bearbeitet von Bernward Helfer und Karl-Ernst Lupprian

In elektronischer Form veröffentlicht auf der Internetseite der Generaldirektion der Staatlichen Archive Bayerns (<http://www.gda.bayern.de/digeinl.htm>).

Copyright (c) 2004 Bernward Helfer und Karl-Ernst Lupprian

Dieses Werk darf frei heruntergeladen, gespeichert, ausgedruckt und weitergegeben werden.
Eine weitergehende, insbesondere kommerzielle Nutzung ist nur mit ausdrücklicher Genehmigung der Herausgeber und der Autoren zulässig.

Inhaltsverzeichnis

Einleitung	1
1. JEAN-DANIEL ZELLER: Archivierung elektronischer Dokumente	2
1.1 Was ist ein Format?	2
1.2 Elektronisches Dokumentenformat	2
1.3 Definition der verschiedenen normativen Zustände der Formate	2
1.4 Typologie nach der Benutzung	3
1.5 Weiterführende Informationen	4
1.6 Formate von Text-Dokumenten	4
2. JEAN-MARC COMMENT: Formate grafischer Dokumente	14
2.1 Rastergrafik	14
2.2 Vektorgrafik	14
2.3 Mathematisches Modell	15
3. JEAN-MARC COMMENT: Datenbank-Formate	22
4. JEAN-JACQUES EGGLER: Audiovisuelle Formate	25
4.1 Das Digitale Video und seine Formate	25
4.2 Der digitale Ton und seine Formate	30
5. KARL-ERNST LUPPRIAN: Elektronische Post (E-Mail)	35

Einleitung

Elektronische Unterlagen bestehen technisch gesehen aus Dateien. Eine Datei ist gekennzeichnet durch Name, Umfang und ihr Format. Das Dateiformat beschreibt die technische Struktur der Datei und gibt der auf die Datei zugreifenden Software Anweisungen für die weitere Verarbeitung. Die Kenntnis der Dateiformate und deren Eignung oder Nichteignung für die Archivierung ist eine der Grundvoraussetzungen für einen fachkundigen Umgang des Archivars mit elektronischen Unterlagen.

Die vorliegende Handreichung soll Archivarinnen und Archivaren lediglich Grundlagen vermitteln, um sich in Diskussionen mit IT-Spezialisten sachkundig und kritisch einbringen zu können. Die Handreichung kann daher weder vollständig noch auf dem neuesten Stand sein. Sie versucht zu erklären, was Dateiformate sind, wie man sie einteilt und beschreibt, welche Zwecke sie erfüllen und inwieweit sie archivfähig sind. Für die Kategorien Textdokumente, Grafiken (im weitesten Sinne), Datenbanken, audiovisuelle Dokumente und elektronische Post werden jeweils verbreitete Formate vorgestellt und die hinter ihnen stehenden technischen Gegebenheiten erläutert.

Die Kapitel 1 bis 4 sind ursprünglich in französischer Sprache auf den Webseiten des Schweizerischen Archivarsverbandes im Rahmen eines Workshops am 16.5.2000 veröffentlicht worden.¹ Wir danken den Autoren Jean-Daniel Zeller, Jean-Marc Comment und Jean-Jacques Eggler ganz herzlich für ihr Einverständnis, ihre Texte ins Deutsche zu übersetzen, gegebenenfalls zu ergänzen und in elektronischer Form publizieren zu dürfen. Unseren Kolleginnen und Kollegen im EDV-Ausschuss der Konferenz der Archivreferenten des Bundes und der Länder (ARK), auf dessen Anregung diese Handreichung entstanden ist, danken wir für fruchtbare Diskussionsbeiträge. Der Generaldirektion der Staatlichen Archive Bayerns gilt unser Dank für die Bereitstellung der Handreichung auf ihren Webseiten.

Kommentare und Ergänzungen sind jederzeit willkommen.²

Wiesbaden und München, im August 2004

Bernward Helfer Karl-Ernst Lupprian

¹ Die Originaltexte sind einsehbar unter: http://www.staluzern.ch/vsa/ag_aca/ate_1/ate_1_1.html .

² Bernward Helfer: <B.Helfer@hhstaw.hessen.de>

Karl-Ernst Lupprian: <iuk@gda.bayern.de>

1. Archivierung elektronischer Dokumente

von Jean-Daniel Zeller, leitender Archivar, Universitätskliniken Genf

1.1 Was ist ein Format?

Der Übergang von Dokumenten in Papierform zu elektronischen Dokumenten verursacht Unsicherheiten bezüglich des Begriffs und der Definitionen von Formaten. Für Papierdokumente reichen Angaben zu den physischen Eigenschaften im allgemeinen weitgehend aus, und diese beschränken sich meistens noch auf eine Beschreibung der externen Form des Dokuments. In einer elektronischen Umgebung müssen Formatangaben dagegen auch Informationen über intrinsische Eigenschaften der Dokumente liefern, die vom Leser nicht wahrgenommen werden können, bevor sie von einem Softwareprogramm übersetzt wurden.

Der Vollständigkeit halber sei angemerkt, dass man auch im Zusammenhang mit Datenträgern von Formaten spricht. In diesem Fall handelt es sich meistens um die physische Beschreibung des Datenträgers, gelegentlich aber auch um Angaben zur logischen Organisation der Daten auf diesem Träger vgl. die Tabelle *Digitale Video-Standards* auf S. 27.

1.2 Elektronisches Dokumentenformat

In Anbetracht des vorher Gesagten muss ein elektronisches Format folglich beschreiben:

- den Vorgang (Inhalt: Text, Bild)
- die Wiedergabe (die Art, in der sie kodiert wird)
- ihre Neugruppierung in einem Dokument (logische Neugruppierung in einer Einheit)

Vermerk: durch einen Link kann die Entsprechung 1 Dokument = 1 Datei aufgehoben werden

- ihre Kompression (wenn diese Funktionalität sich als notwendig erweist)

Die elektronischen Dokumentenformate werden meistens durch die Erweiterung identifiziert, die auf den eigentlichen Dateinamen folgt. Diese Erweiterung des Dateinamens zeigt der Software an, zu welcher Art von Format das Dokument gehört, damit sie es dekodieren und richtig darstellen kann.

1.3 Definition der verschiedenen normativen Zustände der Formate

Zu Beginn des IT-Zeitalters hat die starke Interdependenz zwischen Software und Hardware jeden Softwarelieferanten dazu veranlasst, seine eigenen Formate zu definieren. Die Dissoziation von Hardware und Software hat einen Normungsbedarf bewirkt, zuerst national dann international. Diese Normung, die ein ziemlich langsamer Vorgang ist, wurde von den Herstellern auf zweierlei Art bewerkstelligt:

- durch Bildung eines Konsortiums, das ein Format definiert, bevor das Produkt auf den Markt gebracht wird
(Beispiel: die CD-ROM oder die DVD)
- durch die kostenlose oder sehr preiswerte öffentliche Zurverfügungstellung einer Technologie durch einen Hersteller; die schnelle Verbreitung eines Produktes verursacht eine Normung, indem sie Wettbewerber davon abhält, in einen schon weitgehend besetzten Markt zu investieren (Beispiel: die Musikkassette, die Software Acrobat Reader von Adobe; Gegenbeispiel: Netscape Navigator).

Man kann also bis zum heutigen Tag die nachstehenden Arten von Normung unterscheiden, die vom Einzelfall (keinerlei Norm) bis zum Allgemeinsten (internationale Norm) gehen. In der Praxis gehen die Formate meistens von einer Stufe zu einer anderen über, je nach ihrer Marktverbreitung. Ein proprietäres Format kann sich durch bestimmte Erweiterungen über einen Handelsstandard bis hin zur internationalen Norm entwickeln. (Vermerk: in der französisch sprechenden Welt hat man sich angewöhnt, „Norm“ eine anerkannte internationale

Norm zu nennen und „Standard“ eine Handelsnorm. In den angelsächsischen Ländern ist Standard = Norm ungeachtet des Status).

a) Proprietäres Format (Eigentümerformat)

Format, das für eine Anwendung eines bestimmten Herstellers spezifisch ist.

Unter archivischen Gesichtspunkten sind Formate dieser Art abzulehnen, da sie keine Dauerhaftigkeitsgarantie bieten.

b) de facto Standard

Format, das für einen Hersteller spezifisch ist, das weitverbreitet ist und/oder von vielen anderen Herstellern übernommen wurde, so dass es als eine Tatsachennorm angesehen wird, ohne aber die Dauerhaftigkeit einer institutionellen Norm zu haben.

Unter archivischen Gesichtspunkten können Formate dieser Art als vorläufiges Aufbewahrungsformat (auf mittlere Frist) benutzt werden, da sie vielen Benutzern zugänglich gemacht werden können.

c) Handelsstandard

Format, das durch ein Herstellerkonsortium bereits vor der Markteinführung ausgearbeitet wurde. Manche Konsortien wählen diese Vorgehensweise aus folgenden Gründen:

- da sie schneller ist als die Ausarbeitung einer internationalen Norm
- da sie vermeiden hilft, dass kostspielige Investitionen verlustig gehen, sei es wegen mangelnder Marktakzeptanz, sei es durch aggressive Wettbewerber
- um später vorzuschlagen, dass ihr Handelsstandard als Norm akzeptiert wird.

Unter archivischen Gesichtspunkten können Formate dieser Art als vorläufiges Aufbewahrungsformat (auf mittlere Frist) benutzt werden, da sie vielen Benutzern zugänglich gemacht werden können.

d) Internationale Norm

Format, das im Rahmen der ISO (Internationale Standard Organisation) oder anderer Fachgremien (im Rahmen des Internet ist es z.B. das W3C, das diese Rolle wahrnimmt) entwickelt wurde. Die in diesem Rahmen entstandenen internationalen Normen bieten eine gewisse Dauerhaftigkeitsgarantie, da sie durch die zuständigen nationalen und internationalen Organisationen unterstützt werden. Leider geht das Verfahren einer Normentwicklung sehr langsam vor sich, was manche Marktteilnehmer veranlasst, den Vorgang zu verkürzen, indem sie neue, nicht genormte, aber manchmal leistungsstärkere Produkte anbieten.

1.4 Typologie nach der Benutzung

Im Archivrahmen werden bestimmte Formate für Veröffentlichungszwecke, nicht aber für eine langfristige Archivierung brauchbar sein. Im Einzelnen unterscheidet man:

Ursprungsformat (oder Erwerbsformat)

Format, in dem das Dokument produziert worden ist. Das kann gleichzeitig das Austauschformat oder das Verbreitungsformat, aber selten das Konservierungsformat sein.

Beispiel: alle Eigentümerformate

Austauschformat

Format, in das das Ursprungsformat übertragen worden ist, um darin die Bearbeitung (Lesen und Schreiben) durch eine andere Software zu ermöglichen.

Beispiel: RTF, PDF.

Verbreitungsformat (oder Anzeigeformat)

Format, in das das Ursprungsformat übertragen worden ist, um die Anzeige (Lesen) durch eine andere Software zu ermöglichen.

Beispiel: HTML, Postscript.

Konservierungs- oder Archivierungsformat

Format, in das die Ursprungs-Datei (eventuell nach Umsetzen in ein Austauschformat), übertragen worden ist, um damit die langfristige Archivierung zu ermöglichen.

Beispiel: es gibt momentan kein Format, das speziell für diese Funktion entwickelt wurde, aber SGML oder XML bzw. PDF/A sind mögliche Kandidaten.

1.5 Weiterführende Informationen

Man kann für diesen Bereich mehrere besonders einschlägige Web-Sites empfehlen:

Als Einstieg in die Fachterminologie ist geeignet: <http://www.glossar.de> .

Für die Suche nach einem bestimmten Dateiformat sind hilfreich:

<http://www.wotsit.org>, <http://bton.com/tb17/formats.html> und <http://www.whatis.com>

Für eine Liste der einschlägigen Normen, ständig auf den neuesten Stand gebracht, führt die von der DG XIII der Europäischen Gemeinschaft beauftragte Gruppe IST (Information Science Technology) eine durch Führer und Aktualisierungen auf dem Gebiet vervollständigte bemerkenswerte Arbeit durch, deren Lektüre Pflicht ist (in Englisch):

<http://158.169.50.95:10080/oii/en/guides.html> oder

<http://www.diffuse.org/diffuse.html>

[Anmerkung der Bearbeiter: Leider sind diese Links z.Z. nicht verfügbar]

Links auf Sites, die bestimmte Formate ausführlich beschreiben, werden in den betreffenden Kapiteln angegeben.

1.6 Formate von Text-Dokumenten

Im Folgenden wird für jedes häufig benutzte Format eine Beschreibung geliefert. Sie gibt an:

- den normativen Status des Formats und seinen Typ (nach der nachstehenden Typologie)
- seinen Ursprung, seine Ziele, seine Hauptfunktionen.

Ein zweiter Abschnitt stellt die Vorteile/Nachteile unter Informatik- und Archivgesichtspunkten dar. Schließlich gibt ein Abschnitt die Referenzen oder die Web-Adressen an, die Zusatzinformationen liefern.

Die Mehrzahl der technischen Daten über diese Formate stammt von der empfehlenswerten Site der Open Exchange-Information - OII der DG XIII der Europäischen Gemeinschaften (derzeit verfügbar auf der Site <http://www.diffuse.org/diffuse.html>). Diese Site enthält Normenlisten und nach Typ zusammengefasste Spezifikationen (Standard List) sowie eine Reihe von thematischen Führern (OII Führer). Schließlich bietet sie eine Beobachtung der Aktivitäten in diesen Bereichen an (OII Fora List, Conference Reports, Monthly Reports, OII FAQ usw.). Der OII Führer *Archiving* (http://www2.echo.lu/oii/en/archives_guide.html), bildet den unumgänglichen Einführungstext in diesen Bereich. [Anmerkung der Bearbeiter: Leider sind diese Links z.Z. nicht verfügbar]

Elemente, aus denen sich elektronische Dokumente zusammensetzen

In der Praxis kann sich ein Format auf ein oder mehrere Formate niedrigerer oder angrenzender Ebenen stützen. Man nimmt im allgemeinen an, dass ein Dokument sich aus 4 Hauptele-

menten zusammensetzt, die durch ein einheitliches Format oder durch Formatgruppen definiert werden können:

- der Inhalt (im vorliegenden Fall der rohe Text)
- die Struktur (logische Organisation des Textes)
- der Zusammenhang (Meta-Daten)
- die Darstellung (layout)

Je nach Format kann die Struktur gemeinsam mit dem Inhalt (embedded layout) oder getrennt von diesem bewahrt werden.

Der Kontext wird in den sogenannten Metadaten bewahrt, deren einschlägigstes Beispiel bis zum heutigen Tag in der Archivwelt die EAD (Encoded Archival Description) ist. Außer Daten, die jedem Dokument gemeinsam sind (Autor, Datum, Empfänger), entwickelt jeder Tätigkeitsbereich aktuell seine eigenen Metadaten entsprechend dem Gebrauch, den er von seinen Dokumenten machen will. Die Normung in diesem Bereich kann also nur sektoriell sein und wird höchstens nach einer langen Reifezeit zu einer universellen Norm führen. Das impliziert, dass Dokumente, die gemeinsam in verschiedenen Tätigkeitsbereichen benutzt wurden, mehrere Konglomerate von Metadaten gemäß den verschiedenen Anwendungszwecken erhalten können.

Typologie der elektronischen Dokumentenformate (Texte)

Aus Gründen der Bequemlichkeit unterscheidet man im Augenblick die folgenden Arten von Formaten, in der Reihenfolge wachsender Komplexität:

- a) Zeichenformat oder Kodierung (in Englisch: character set)
es ist das Ausgangsniveau, das die Entsprechung zwischen der digitalen Kodierung der Maschine und den Zeichen des Alphabets (unter Einschluss der diakritischen Zeichen und der Zahlen) definiert.
- b) Format der Text-Bearbeitung
dieses Niveau ist dem Zeichenformat übergeordnet, indem es ihm Angaben zur Paginierung, zum Seitenumbruch, eventuell zum Inhaltsverzeichnis hinzufügt. In manchen Fällen umfasst es direkt Eigenschaften des Anzeigeformats (z.B. MS-Word), in anderen Fällen nicht (z.B. SGML).
- c) Austauschformat (oder Übertragungsformat),
die Vielfalt proprietärer Textformate hat die Entwickler veranlasst, Austauschformate zu definieren (auch Exportformate genannt), die als gemeinsame Plattform für Dateien/Dokumente dienen, die durch unterschiedliche Software entstehen, und die die Wiederverwendung des Dokuments durch eine andere als die Ursprungs-Software erlauben.
- d) Anzeigeformat
aus den selben Gründen der Kompatibilität und der Übertragung sind Anzeigeformate (in Englisch: layout), entstanden. Diese unterscheiden zwischen Basisdokumenten, welche die Rohdaten enthalten und Angaben zur Anzeige und zum Umbruch (stylesheet in Englisch.)

Die nachstehende Tabelle fasst die gegenseitigen Beziehungen zwischen verschiedenen Formaten zusammen, die auf Grund ihrer Funktionalitäten beschrieben wurden. Die Format-Typen werden in *Kursivschrift* angegeben. Die genormten Formate (im strengen Sinn), sind darin **fett** enthalten. Die SGML verwandten Formate sind grau unterlegt

<i>Zeichenformat</i>						
ASCII - ISO 646 (.txt) - ISO 10646 (UCS)						
<i>Text-Bearbeitungs-Format</i>		<i>Austauschformat</i>			<i>Anzeigeformat</i>	
WordPerfect (.wpd)	Word (.doc)	RTF	PostScript	SGML	HTML (.html)	HyTime
			PDF	XML		
				DSSSL	XSL	CSS2
<i>Layout-Formate</i>						
	Proprietär	---	----	z.B.: EAD		
<i>Meta-Daten</i>						
Dublin Core						

Format:	ASCII
	American Standard Code for Information Interchange
Status	Nationale Norm
Typ	Zeichenformat, eigentlich Zeichenkodierung
Geschichte	Entwickelt von American National Standard Institute (ANSI), bildet der ASCII-Code momentan die amerikanische Version des Codes ISO 646. Hat das konkurrierende proprietäre Format EBCDIC von IBM abgelöst.
Funktionalität:	Dieser Code definiert 94 Zeichen (Buchstaben, Zahlen, diakritische und mathematische Zeichen), und das Spatium, die in Dokumenten in englischer Sprache benötigt werden (ohne Akzente).
Informatikvorteile:	Norm, die universell benutzt wurde, um Zeichen darzustellen.
Informatiknachteile:	Umfasst weder die Informationen über die Formate betonter Zeichen (Akzente) noch Informationen über Auszeichnungen (fett, kursiv usw.).
Archivvorteile:	Auf Grund seiner weiten Verbreitung mit jeder beliebigen Software lesbar.
Archivnachteil:	Brauchbar nur für Sprachen ohne betonte Zeichen. Enthält keine Informationen über das Layout
Mehr Informationen	Liste des Codes: http://www.dkuug.dk/i18n/charmmaps

Format:	ISO 646
	ISO 7-Bit coded character set for Information interchange
Status	Internationale Norm
Typ	Zeichenformat, eigentlich Zeichenkodierung
Geschichte	Basiert auf dem ASCII-Code. Er vervollständigt diesen, indem er die über die Shift-Tasten verfügbaren Akzente umfasst.
Funktionalität:	Dieser Code definiert 127 Zeichen (Buchstaben, Zahlen, diakritische und mathematische Zeichen), und das Spatium. Er ist brauchbar für Dokumente in europäischen Sprachen (einige Codierungen stellen unterschiedliche Akzente in den diversen Sprachen dar)
Informatikvorteile:	Norm, die universell benutzt wurde, um Zeichen darzustellen.
Informatiknachteile:	Enthält keine Informationen über Auszeichnungen (fett, kursiv usw.).
Archivvorteile:	Auf Grund seiner weiten Verbreitung mit jeder beliebigen Software lesbar..
Archivnachteile:	Keine Information über das Layout. Unterschiedliche Kodierung der Akzente je nach Sprache.
Mehr Informationen	ISO/IEC646:1991

Format:	ISO 8859
	ISO 8-Bit coded character set for Information interchange
Status	Internationale Norm
Typ	Zeichenformat, eigentlich Zeichenkodierung
Geschichte	Basiert auf dem 7-Bit-ASCII-Code und erweitert diesen um ein Bit.
Funktionalität:	Dieser Code definiert 254 Zeichen, deren Gesamtheit auch als Codepage bezeichnet wird. Es gibt mehr als ein Dutzend verschiedene Codepages, deren erste Hälfte den 7-Bit ASCII-Code enthält, während die zweite Hälfte vor allem für nationale Besonderheiten verwendet wird. Für die westeuropäischen Sprachen dient ISO 8859-1 (auch ISO Latin-1 genannt).
Informatikvorteile:	Norm, die universell benutzt wurde, um Zeichen darzustellen.
Informatiknachteile:	Enthält keine Informationen über Auszeichnungen (fett, kursiv usw.).
Archivvorteile:	Auf Grund seiner weiten Verbreitung mit jeder beliebigen Software lesbar..
Archivnachteile:	Keine Information über das Layout.
Mehr Informationen	ISO/IEC:8859

Format:	ISO 10646
	Universal Multiple-Octet-Byte Coded Character Set (UCS oder UNICO-DE)
Status	Internationale Norm
Typ	Zeichenformat, eigentlich Zeichenkodierung
Geschichte	Code, der die meisten der früheren Zeichenkodes integriert, indem er ein Zeichen mit 16 oder 32 Bit darstellt.
Funktionalität:	Dieser Code definiert 127 Zeichen, die den "BASIC Multilingual Plan" (BMP) bilden, der die Spezifizierungen von ISO 646 wieder aufnimmt. Er wird durch die Spezifizierungen der ISO/IEC 8859-1 (europäische Sprachen) vervollständigt, sowie durch nicht-europäische Zeichen und eine Reihe graphischer Symbole. Der Gesamtzeichenvorrat umfasst 65.536 Zeichen.
Informatikvorteile:	Zur Zeit erschöpfendste Norm um Zeichen darzustellen, ist aber wegen mangelnder Implementierung auf Standardrechnern noch wenig verbreitet.
Informatiknachteile:	Enthält keine Informationen über Auszeichnungen (fett, kursiv usw.).
Archivvorteile:	Auf Grund seiner weiten Verbreitung mit jeder beliebigen Software lesbar.
Archivnachteile:	Keine Information über das Layout.
Mehr Informationen	http://www.dkuug.dk/i18n/ISO_10646 .- The Unicode Standard, Version 3.0, Addison Wesley Longman, 2000 (ISBN: 0-201-61633-5)

Format:	Winword (Ms-Word)
Status	Proprietäres Format (Microsoft)
Typ	Textformat
Geschichte	Textverarbeitungssoftware, die ursprünglich durch Microsoft in ASCII-Code entwickelt wurde, und dann, in immer neuen Versionen, ständig erweitert wurde.
Funktionalität:	Word als Urtyp der Software einer Textverarbeitung für Geschäftstexte erlaubt die Gestaltung mehrseitiger Dokumente, einschließlich <ul style="list-style-type: none"> - Texterfassung (ASCII, mit Microsoft-eigenen Erweiterungen) - Auszeichnungen (RTF) - Layout <p>Word erlaubt, Vorlagen zu definieren. Dabei handelt sich um Modelle, die vom einzelnen Benutzer für seinen Gebrauch entworfen werden. Diese Vorlagen sind nicht Gegenstand einer genormten Beschreibung.</p>
Informatikvorteile:	Word ist eine weitverbreitete Software, was eine gewisse Dauerhaftigkeit gewährleistet.
Informatiknachteile:	Als Handelsformat sind Aktualisierung häufig (alle 3-5 Jahre). Es ist (bisher) aufwärtskompatibel. Abwärtskompatibilität ist nicht gewährleistet.
Archivvorteile:	Keine, mit Ausnahme der weiten Verbreitung
Archivnachteile:	Proprietäres Format, dessen Versionen etwa alle 3 Jahre wechseln. Die manchmal künstliche Schaffung neuer Versionen macht ihre langfristige Benutzung gefährlich

Mehr Informationen	
--------------------	--

Format:	HTML
	HyperText Markup Language
Status	Internationale Norm
Typ	Anzeigeformat, eigentlich eine Auszeichnungssprache (Markup Language)
Geschichte	HTML ist eine Teilmenge von SGML, besonders ausgearbeitet, um Dokumente im Rahmen des World Wide Web anzuzeigen. (Schaffung 1990). Sie wird durch das W3C (WWW Konsortium) verwaltet, und es gibt ein ISO Normierungs-Projekt zu diesem Thema. Mit der Version 4 wurde HTML abgeschlossen; es wird nunmehr die Verwendung des Nachfolgers XHTML (eXtended HTML) empfohlen, dessen Syntax strenger ist.
Funktionalität:	HTML ist geeignet zur Anzeige einfach strukturierter und relativ kurzer Texte. Bilder können eingebunden werden.
Informatikvorteile:	Von der materiellen Infrastruktur unabhängige universelle Sprache, deren Konzept (Benutzung von Markierungszeichen) eine große Flexibilität erlaubt. Die neueren Versionen der meisten Textverarbeitungsprogramme enthalten Funktionen zur direkten Umsetzung nach HTML
Informatiknachteile:	In der Praxis erfolgt die Anzeige langer und komplex strukturierter Texte immer häufiger im PDF-Format.
Archivvorteile:	Neben der weiten Verbreitung ist vorteilhaft, dass HTML-Dokumente reine Textdateien mit 7-Bit-ASCII-Kodierung darstellen, deren Inhalt zur Not auch ohne Webbrowser gelesen werden kann.
Archivnachteile:	HTML ist ein Anzeigeformat, was eine Anpassung an die Umgebung, in der die Visualisierung erfolgt (Webbrowser), impliziert. Das hat zur Folge, dass ein HTML-Dokument ein unterschiedliches Aussehen haben kann, es sei denn, der Schöpfer des Dokuments habe zwingende Markierungszeichen angebracht, was aber im Gegensatz zur Flexibilität der Sprache stünde und daher meistens nicht getan wird.
Mehr Informationen	http://www.w3.org/MarkUp/Activity.html .- Stefan Münz, SelfHTML. Sowohl als Buch als auch online verfügbar: http://aktuell.de.selfhtml.org/extras/download.shtml .

Format:	HyTime
	Hypermedia/Time-based Structuring Language
Status	Proprietäres Format (Adobe) in Entwicklung zur Norm
Typ	Austauschformat, eigentlich strukturierende Sprache
Geschichte	vorgestellt als Eigentümerformat, adaptiert als Norm im Jahre 1992 (ISO 10744:1997)
Funktionalität:	HyTime ist eine auf der SGML Struktur basierende Sprache, die die Einbeziehung textueller und dynamischer (Töne, Videosequenzen usw.) Hypermedia-Dateien erlaubt. Diese Möglichkeit nimmt ihm dagegen die Flexibilität, die SGML eigen ist, denn die audiovisuellen Komponenten werden nach einer HyTime eigenen Semantik interpretiert. Diese Objekte werden nicht in das HyTime-Dokument eingebettet. Vielmehr liefert es Werk-

	zeuge, die den Zugang zu diesen Dokumenten ermöglichen.
Informatikvorteile:	Ermöglicht den Zugang zu Hypermedia-Dokumenten, ohne die Struktur der Quelldaten kennen zu müssen
Informatiknachteile:	Momentan bieten die HyTime-Software-Erzeugnisse nicht alle Funktionalitäten der Norm an, sondern konzentrieren sich auf eine einzelne Art von Daten nach den Bedürfnissen spezifischer Anwendungen (einzig bekanntes kommerzielles Produkt "HyMinder" von TechnoTeacher).
Archivvorteile:	Keine
Archivnachteile:	Neue Norm mit komplexer Verwaltung, die auf den Austausch dynamischer Daten, aber nicht auf deren Archivierung ausgerichtet wurde.
Mehr Informationen	HyTime scheint keine Verbreitung beschieden zu sein.

Format:	PDF
	Portable Document Format
Status	Ursprünglich proprietäres Format (Adobe), inzwischen freigegeben (nur die Spezifikationen unterliegen dem Copyright von Adobe / de Facto-Format
Typ	Austauschformat
Geschichte	Format, das von Adobe geschaffen wurde, um den einfachen Austausch umfänglicher Dokumente im World Wide Web zu erlauben. Derivat von PostScript.
Funktionalität:	Ein PDF-Dokument wird durch die Acrobat-Software (kostenpflichtig) erzeugt und kann durch die Software Acrobat Reader (kostenlos) gelesen werden (und kopiert oder gedruckt, wenn der Autor es in der Parameterdarstellung bei der Schaffung des Dokuments erlaubt hat). Inzwischen gibt es auch Software anderer Hersteller, die eine Erzeugung von PDF-Dokumenten erlaubt. PDF-Dokumente erhalten das ursprüngliche Layout eines Textes.
Informatikvorteile:	In Anbetracht der kostenlosen Verbreitung von Acrobat Reader wird das PDF-Format breit genutzt und bildet einen de facto-Standard für den Dokumentenaustausch im Internet
Informatiknachteile:	
Archivvorteile:	Praktisches Kommunikationsformat für umfängliche Dokumente, das auch Graphiken einschließt, die aber nicht immer richtig wiedergegeben werden. Eine für Archivierungszwecke besonders geeignete Variante (PDF/A) ist in Entwicklung.
Archivnachteile:	PDF-Dateien können aktive Inhalte bergen, was jedoch für PDF/A vermieden werden soll.
Mehr Informationen	Portable Document Format Reference Manual, Addison Wesley Longman 1996 ISBN 0.201.62628 4.- PDF als Archivierungsstandard. White Paper 2003, Adobe Systems, zu finden unter http://www.adobe.com .

Format:	RTF
	Rich Text Format
Status	Proprietäres Format (Microsoft)
Typ	Austauschformat
Geschichte	Geschaffen durch Microsoft, um die Weiterverwendung der in den unterschiedlichen Textverarbeitungen des Hauses (Word-Versionen) entstandenen Texte mit ihren Auszeichnungen und ihrem Layout zu ermöglichen.
Funktionalität:	Dieses Format erlaubt die Umsetzung von Dokumenten, die durch den Großteil der marktgängigen Textverarbeitungsprogramme erzeugt wurden, in beide Richtungen (die Umsetzung ist allerdings nur für die Eigenschaften gewährleistet, die allen diesen Textverarbeitungen gemeinsam sind, was manchmal Überraschungen bereit hält).
Informatikvorteile:	Da RTF die am weitesten verbreiteten Layout-Charakteristika eines Dokumentes nutzt, stellt es eine Methode des Dokumentenaustausches dar, die weitgehend in Konkurrenz zu PDF tritt.
Informatiknachteile:	Da Microsoft bei jeder Versionsänderung von Word auch die Version von RTF wechselte, gibt es keine Garantie, dass die übrigen Lieferanten diesen Aktualisierungen korrekt folgen.
Archivvorteile:	
Archivnachteile:	Proprietäres Format, ständiger Versionswechsel.
Mehr Informationen	Word for Windows Technical Reference Manual, Microsoft Press, mehrere Auflagen

Format:	SGML
	Standardized General Markup Language
Status	Internationale Norm (ISO 8879:1986)
Typ	Austauschformat, eigentlich eine Metasprache
Geschichte	Entwickelt im Rahmen der ISO, basiert auf einer objektorientierten Methode, definiert diese Norm eine Semantik, um die Struktur von Dokumenten zu beschreiben, und eine Syntax, um diese Beschreibung (DTD) formell zu kodieren. HTML ist davon eine Teilmenge, die wahrscheinlich mit der Zeit durch XML ersetzt wird.
Funktionalität:	SGML erlaubt, die Dokumente mit ihrer Struktur aufzubewahren (Titel, Kopf, Kapitel, Paragraph usw.) aber ohne ihr Layout. Die Struktur wird durch eine DTD (Document Type Definition) definiert. Man kann das Layout durch ein "stylesheet" vervollständigen, das durch die DSSSL (Document Style Semantic and Specification Language) definiert wird. Eine DTD kann für den Gebrauch nur einer Institution definiert werden, aber ihr Hauptwert besteht darin, durch eine Nutzergemeinschaft angewandt zu werden, wie dies beispielsweise bei der Nutzung der Encoded Archival Description (EAD) geschieht.
Informatikvorteile:	
Informatiknachteile:	SGML erfordert die Existenz einer DTD für jeden produzierten Dokumententyp, was eine funktionierende Kontrolle seitens der produzierenden Stelle voraussetzt. Wegen der Komplexität und des Umfangs der Sprache gibt es nur wenige Softwareprodukte, die eine komfortable Dokumentenerstel-

	lung mit SGML unterstützen.
Archivvorteile:	In SGML werden die Dokumente mit ihrer logischen Struktur gespeichert. Man kann sie also nach ihrer DTD oder nach irgendeinem benutzten Markierungszeichen sortieren, was erlaubt, je nach Bedarf sehr hochentwickelte Rechercheinstrumente zu generieren. Ferner sind SGML-Dokumente reine, mit 7-Bit-ASCII kodierte Textdateien, die zur Not mit jedem einfachen Texteditor gelesen werden können.
Archivnachteile:	
Mehr Informationen	http://xml.coverpages.org/sgml.html

Format:	XML
	Extended Markup Language
Status	Internationale Norm (W3C)
Typ	Austauschformat, eigentlich Auszeichnungssprache
Geschichte	SGML-Teilmenge, konstruiert, um komplexe Dokumente im Internet übermitteln zu können, ohne zwingend eine Dokument Type Definition (DTD) verwenden zu müssen. Überschaubarer und leichter zu handhaben als SGML. Verwaltet durch das W3C.
Funktionalität:	Anstelle einer DTD können auch sogenannte XML-Schemata verwendet werden, um Dokumente zu kodieren. Schemata erlauben weitergehende Validierungen (parsing), als dies eine DTD ermöglicht. Ein Layout kann mit Cascading Stylesheets (CSS) oder der Sprache XSL (Extended Stylesheet Language) erstellt werden.
Informatikvorteile:	Webbrowser der neueren Generationen können XML-Dokumente verarbeiten. Es gibt bereits zahlreiche Softwareprodukte, um solche Dokumente zu erstellen und zu bearbeiten.
Informatiknachteile:	
Archivvorteile:	XML-Dokumente sind reine, mit 7-Bit-ASCII kodierte Textdateien, die zur Not mit jedem einfachen Texteditor gelesen werden können.
Archivnachteile:	
Mehr Informationen	www.w3.org/XML/

Format:	PostScript
Status	Proprietäres Format (Adobe)
Typ	Seitenbeschreibungformat
Geschichte	Format, von Adobe entwickelt (1985), um Dokumente, die Graphiken als Raster oder in Vektordarstellung enthalten, an irgendeine Druckerart versenden zu können.
Funktionalität:	PostScript ist ein indirektes Austauschformat, das hauptsächlich für die Übermittlung von Druckfunktionalitäten genutzt wird. Für den eigentlichen

	Dokumentenaustausch hat Adobe später das PDF-Format entwickelt. In der Unix- und Linux-Welt ist PostScript das Standardformat für den Druck.
Informatikvorteile:	
Informatiknachteile:	
Archivvorteile:	
Archivnachteile:	Proprietäres Format mit begrenzter Funktionalität. Hauptsächlich für graphische Dokumente.
Mehr Informationen	PostScript Language Reference Manuel, Addison Wesley Longman, 1990, ISBN 0.201 18127-4.- http://www.adobe.com

2. Formate grafischer Dokumente

von J.-M. Comment, Schweizerisches Bundesarchiv

2.1 Rastergrafik

Die einfachste Methode, ein bestehendes Bild zu zerlegen, besteht darin, es in einer Zellenmatrix durch ein einfaches Gitternetz zu zerschneiden. Diese Zellen sind meistens Quadrate, die als unteilbar gelten und die durch eine einzige Farbe charakterisiert werden. Man nennt sie *Pixel* vom Englischen *picture element*. Je kleiner die Größe der Pixel, je näher ist das Modell am dargestellten Bild. Die Auflösung des Bildes, die der Anzahl Pixel je Längeneinheit entspricht, wird oft in Punkten per Zoll ausgedrückt (ppp oder dpi - dot per inch - in Englisch). Die Bilder, die diesem Modell entsprechen, werden als Rastergrafik bezeichnet (häufig auch als bitmap, um den gewöhnlich benutzten englischen Begriff wieder aufzunehmen). Dieses Modell läuft darauf hinaus, ein Bild als eine Gesamtheit von Punkten zu betrachten.

Farbcodierung

Art der Kodierung	Bit pro Pixel	Mögliche Werte	Interpretation dieser Werte
Strichzeichnung	1 bit	0 oder 1	0 = schwarz, 1 = weiß
Schwarz und weiß mit Graustufen	8 bit	von 0 bis 255	256 Graustufen, von schwarz (0) bis weiß (255).
Kodierung der VGA-Grafikkarten	4 bit	von 0 bis 15	bit 1 = rot, bit 2 = grün, bit 3 = blau, bit 4 = Intensität, das erlaubt 16 Kombinationen.
Echtfarben (<i>true color</i>)	24 bit (3x8) oder 15 bit (3x5)	3 Werte von 0 bis 255 oder von 0 und 31	Jeder Wert vertritt eine Intensität von rot, grün oder blau. 0 = keine Farbe; 255 = maximale Intensität. Das ergibt 16 Millionen Farben (24 bit oder 32 000 Farben zu 15 bit).
Farbpalette	8 bit	von 0 bis 255	Eine Tabelle namens Palette definiert die Beziehung zwischen diesen 256 Farben und den Farben in der Wirklichkeit; kodiert nach dem RGB-Modell.

2.2 Vektorgrafik

In bestimmten Fällen ist es interessanter, ein Bild mittels komplexerer Objekte, wie Segmente, Kreise oder, ganz allgemein, vordefinierter geometrischer Formen, zu zerlegen. Das Bild wird dann durch ein komplexeres mathematisches Modell beschrieben, das aber ausreicht, um es darzustellen. Man spricht dann von einer Vektorgraphik, und man findet diese Art der Beschreibung in zahlreichen Grafikprogrammen, die es dem Benutzer erlauben, diese geometrischen Gegenstände und nicht nur Punkte zu handhaben.

Newpath

270 360 moveto

0 72 rlineto

```

72 0 rlineto
0 -72 rlineto
closepath
4 setlinewidth
stroke showpage

```

Beispiel für die Darstellung eines Rasters in Postscript zwecks Übergabe an einen Drucker

2.3 Mathematisches Modell

Der Zweig der Mathematik, der sich mit den Fragen der Kompression befasst, befindet sich in voller Entwicklung. Er stützt sich auf mathematische Modelle und Methoden, die die in der Information enthaltenen Redundanzen eliminieren und eine sehr hohe Datenkompression bei gleichzeitiger Minimierung der Informationsverluste ermöglichen. Dies ist der Fall bei fraktalen Bildern oder Wavelets.

Grafikprogramme

Bitmap

- auf Mac und PC: Photoshop;
- auf PC: Bild Pals, Photo Styler, Corel Photo Paint;
- auf Mac: Superpaint.

Vektorgrafik

- auf Mac und PC: Illustrator;
- auf PC: Microsoft Draw, Corel Draw, Micrografx Draw, Autocad;
- auf Mac: Draw, FreeHand.

Digitalisierung

Scanner, Photo-CD, digitale Kamera, Speicherplatz.

Formate von Bildern

Verlustbehaftete Kompression

Diese Formate werden vorzugsweise für die Integration der Bilder in das Endprodukt benutzt.

Die JPEG-Norm (Joint Photographic Expert Group)

JPEG ist ein von der ISO normierter Algorithmus, der das Kodieren und die Kompression fester Bilder vom Matrizen typ regelt. Der Grundsatz beruht auf der Aufgliederung des Bildes in Matrizen von 8 x 8 Pixeln, der Anwendung einer Diskreten Cosinus-Transformation (DCT), um die Anzahl der Parameter des Bildes, die Messung der Parameter zu reduzieren, und einer Kompression nach Huffman. Diese Kompression bietet progressive abstuftende Verdichtungsverhältnisse an (1:10 ist ein Durchschnitt). Die Kompression wird entweder auf dem Softwareweg oder durch spezifische Hardware durchgeführt.

IFS-Fraktale

Iterated Function System = Systeme von Funktionen die "iteriert" werden, d.h. immer wieder ausgeführt werden. IFS-Fraktale beruhen auf der Idee, dass ein Bild als eine Gesamtheit von Motiven beschrieben werden kann, die sich wiederholen und sich ähnelnden geometrischen Umwandlungen unterliegen. Ein fraktales Bild wird also durch die Gesamtheit dieser Umwandlungen beschrieben und nicht mehr Punkt für Punkt. Diese Verfahrensweise führt dazu, dass das Kodieren eines Bildes von seiner Größe unabhängig wird, und bietet starke Verdichtungsverhältnisse an (von 1:10 bis zu 1:30 im Durchschnitt). Die Kompression wird entweder auf dem Softwareweg oder hardwareseitig durchgeführt. Der entsprechende Code ist stark asymmetrisch. Die Dekompression ist extrem schnell (Sekunde), was diese Art von Bildern sehr interessant macht für ein Editierwerkzeug. Die Kompression ist andererseits langsamer (von 30 s bis zu 5 min, je nach Hardware). Es wird daher empfohlen, bei der Produktion einer großen Anzahl von fraktalen Bildern spezifische Kompressionskarten einzusetzen. Die Klarheit des Bildes geht progressiv zurück, wenn das Verdichtungsverhältnis steigt, aber es gibt keine Mosaikwirkung. Dieses proprietäre Format wird momentan vermarktet und verbreitet sich. Die Nutzungsrechte sind abhängig vom Erwerb der Software und der Hardware, die für die Kompression notwendig ist.

Wavelet-Bilder

Wavelets, das Ergebnis verschiedener Forschung des HARC (Houston Advanced Research Center), werden zur Zeit diskutiert. Die Verdichtungsverhältnisse sind auch hier spektakulär (von 1:10 bis zu 1:50 im Durchschnitt). Die Qualität kann noch nicht beurteilt werden. Es bleibt festzuhalten, dass die durch REA vermarktete Software ImMix diesen Algorithmus benutzt.

Zusammenfassende Tabelle der wichtigsten Bildformate

Format	Typ	Anmerkungen
TIFF <i>Tagged Image File Format</i> Macintosh: ja PC: Dateierweiterung: .TIF	Rastergrafik Verlustfreie Komprimierung möglich.	Klassisches Bildformat Alle Typen von Rastergrafiken: schwarz/weiß und Farbe.
BMP <i>Bitmap format</i> Macintosh: wird erkannt PC: Dateierweiterung: .BMP	Rastergrafik Ohne Komprimierung.	Standardformat in Windows Alle Typen von Pixelbildern: schwarz/weiß und Farbe.
PICT Macintosh: ja PC : wird erkannt	Rastergrafik Ohne Komprimierung.	Standardformat für Mac Alle Typen von Pixelbildern: schwarz/weiß und Farbe.
PaintBrush Macintosh: wird erkannt PC: Dateierweiterung: .PCX	Rastergrafik Verlustfreie Komprimierung.	Proprietäres Format der Software PaintBrush Schlecht bei mehr als 256 Farben
GIF <i>Graphics Interchange Format</i> Macintosh: ja PC : Dateierweiterung: .GIF	Rastergrafik Starke verlustfreie Komprimierung.	Komprimierendes Bildformat, das sich für schnelle Datenübertragung eignet. Ist auf 256 Farben beschränkt.

JPEG <i>Joint Photographic Expert Group</i> Macintosh: ja PC: Dateierweiterung: .JPG	Rastergrafik Internationale Norm. Verlustbehaftete Komprimierung.	Alle Typen von Rastergrafiken: schwarz/weiß und Farbe. Entsprechend der Kompressionsrate (1:10, 1:30, etc.) sind mehrere Qualitätsstufen möglich. (1: 20 ist eine akzeptable Rate)
Photo CD von Kodak Macintosh : ja PC: Dateierweiterung: .PCD	Rastergrafik Format für Photo CD Leichte verlustfreie Komprimierung.	24 Bit Rastergrafik in 5 Auflösungen
Autocad Macintosh: ? PC: Dateierweiterung: .DWG	Vektorgrafik	Standardformat unter Windows zur Wiedergabe von Plänen.
Encapsulated Postscript Macintosh: ja PC: Dateierweiterung: .EPS	Vektorgrafik	Standardbeschreibungsformat für PostScript-Drucker
Fractal Macintosh: ja PC: Dateierweiterung: .FIF	Mathematisches Modell. Verlustbehaftete Komprimierung	Proprietäres Format (in Entwicklung). Mehrere Qualitätsstufen in Abhängigkeit von der Kompressionsrate (von 1:10 bis 1:50) möglich.

Konservierung

Speicherung und Archivierung

Hierfür ist von Formaten mit verlustbehafteter Komprimierung abzuraten, (Ausnahmen sollten nur in ganz besonderen Fällen gemacht werden)

auf Photo-CD: Format Photo-CD. Diese Wahl wird dringend empfohlen, wenn die Bilder für einen Druck guter Qualität auf Papier vorgesehen sind.

auf CD-ROM, Festplatte, usw. im Format: Photo-CD

Rastergrafik (Bitmap): 16 Millionen Farben ohne verlustbehaftete Kompression (Standards wie BMP, PICT, TIFF usw.) mit ausreichenden Definitionen, was aber zu riesigem Speicherbedarf führt,

Vektorgrafik: (Standards wie EPS, WMF, ECW)

Speicherbedarf komprimierter Bilder:

das durch ein Bild besetzte Volumen 640 x 480 in 16 Millionen Farben:

Format	Bitmap BMP	JPEG	Fraktal
Verlustbehaftete Komprimierung	ohne	von 1:10 bis 1:20	von 1:30 bis 1:50
Dateigröße	1 MB	von 100 bis 50 KB	von 30 bis 20 KB

Übergänge

Die Mehrzahl der Bildbearbeitungsprogramme (Photoshop, ImagePals, Photostyler) erlaubt den Wechsel zwischen Dateiformaten, Veränderung der Bildgröße und der Anzahl der Farben usw.. Die Frage des Übergangs zwischen den verschiedenen Standardformaten ist kein wirkliches Problem.

Format	Fax Gruppe 3&4
Status	Internationale Norm (ITU T.4 und T.6)
Typ	Rastergrafik
Geschichte	Kompressionsnormen, die für den Austausch graphischer Daten durch Fernkopierer (Fax) entwickelt und durch das CCITT (technischer Ausschuss der Internationalen Union für Telekommunikation) ausgearbeitet wurden
Funktionalität	Norm für verlustfreie Komprimierung. Gruppe 3 wird für Standard-Fax benutzt und beruht auf einer eindimensionalen Kompression nach Huffman. Gruppe 4 wird für Fax mit hohem Durchsatz, auf der Basis von ISDN, benutzt und beruht auf einer Kompression in den zwei Dimensionen einer Seite.
Informatikvorteile	
Informatiknachteile	
Archivvorteile	Universelle und weitverbreitete Normen.
Archivnachteile	Das TIFF-Format, das den Großteil dieser Funktionalität übernimmt und das ein breiteres Spektrum an Dokumententypen umfasst, sollte ihm vorgezogen werden.
Weitere Informationen	

Format	GIF
	Graphic Interchange Format
Status	Proprietäres Format (Comuserve bzw. Unisys); de facto Standard. Das Patent von Unisys (auf den Kompressionsalgorithmus) lief im Juni 2003 aus, im Juni 2004 liefen auch die europäischen und japanischen Patente aus.
Typ	Austauschformat für Rastergrafik
Geschichte	Entwickelt durch Comuserve, um den Austausch graphischer Daten unabhängig vom benutzten Informatiksystem zu ermöglichen. Comuserve liefert eine begrenzte Freeware-Version.
Funktionalität	GIF speichert die Bilder als eine Pixel-Sequenz in RGB Farbwerten. Jedes Bild beginnt mit einer Kennung (Versionsnummer usw.), einer Bildschirm-Definition (erlaubt die Anpassung des Bildschirms) und einem Farbmaßstab. Die Daten werden mit dem LZW-Algorithmus komprimiert, was einen Datenverlust bei der Vergrößerung verursacht. Die Leichtigkeit der Bedienung und der geringe Speicherbedarf führten dazu, dass GIF sehr stark bei einfachen Graphiken, die in Netzen verbreitet oder versandt werden sollten, eingesetzt wurde. Der LZW-Kompressionsalgorithmus, unterliegt einer Tantiemenzahlung an seinen Eigentümer, die Firma Unisys. Comuserve versuchte daher, einen

	Kompressionsalgorithmus ohne Urheberrecht und ohne Informationsverlust zu entwickeln. Einer der Versuche ist unter dem Namen PNG (Portable Network Graphics) bekannt, aber momentan erscheint JPEG als eine stabilere und sicherere Norm.
Informatikvorteile	Leichte Anwendbarkeit
Informatiknachteile	Der Kompressionsalgorithmus unterliegt derzeit noch Urheberrechten.
Archivvorteile	Weite Verbreitung
Archivnachteile	Praktisches, aber für die dauerhafte Aufbewahrung unbrauchbares Anzeigenformat
Weitere Informationen	Volker König, Lizenzquerelen. Lizenzgebühren für das GIF-Format erhoben. In: c't 3/1995, S.29

Format	JPEG
	Joint Photographic Expert Group
Status	Internationale Norm (ISO/IEC 10918)
Typ	Kompressionsformat für Bilddateien
Geschichte	variables Kompressionsformat, das die Kompression ohne oder mit Informationsverlust erlaubt, gemeinsam entwickelt von ISO und ITU.
Funktionalität	JPEG erlaubt Kompressionen, die vom Verhältnis 2:1 (ohne Informationsverlust) über 8:1 (für den Druck) bis zu 35:1 (für die Anzeige am Bildschirm) reichen.
Informatikvorteile	Gewinne beim Speicherbedarf und bei der Anzeigegeschwindigkeit (Internet) der stark komprimierten Bilder.
Informatiknachteile	
Archivvorteile	Erlaubt die Erhaltung ohne Informationsverlust (aber dann ist der Vorteil der Kompression praktisch nicht vorhanden).
Archivnachteile	Das Verdichtungsverhältnis wird bei der Dateianlage als Maßstab festgelegt, der nicht im Klartext mit den Daten dokumentiert wird. Gutes Kommunikations- oder Austauschformat, aber ungeeignet für die Konservierung.
Weitere Informationen	Soll durch das – völlig anders aufgebaute – Format JPEG2000 ersetzt werden.

Format	Photo-CD
Status	Proprietäres Format (Kodak)
Typ	Anzeige- und Speicherformat
Geschichte	Entwickelt durch Kodak für die Übertragung photographischer Bilder auf den Datenträger CD-ROM. Definiert das Dateiformat und die Organisationsstruktur der CD.
Funktionalität	Die vom Film oder von photographischen Abzügen digitalisierten Bilder, werden in einem „Image Pac“ in 5 unterschiedlichen Auflösungen (100 Photos auf einer CD von 600 MB) gespeichert. Der Vorteil ist der

	direkte Zugang zu einer annehmbaren Auflösung je nach Art der Benutzung (Vignette oder Hochauflösung), der Nachteil ist der durch das „Image Pac“ benötigte Speicherplatz (ungefähr zweimal das Volumen des Bildes in der höchsten Auflösung). Kodak schlägt professionelle Versionen mit höheren Definitionsniveaus vor, die jedoch gleichzeitig die Anzahl der speicherbaren Bilder auf der CD reduzieren.
Informatikvorteile	
Informatiknachteile	
Archivvorteile	Guter Kommunikationsträger in Anbetracht der Handelsverbreitung des Produktes.
Archivnachteile	Schlechter Träger für langfristige Speicherung, weil speicherintensiv und nicht genormt.
Weitere Informationen	http://www.kodak.com/DE/de/consumer/pictureCD/index.shtml

Format	PNG
	Portable Network Graphics
Status	Internationale Norm (W3C)
Typ	Austauschformat (Rastergrafik)
Geschichte	Entwickelt durch das W3C, um das GIF-Format zu ersetzen und um ein von Urheberrechten unabhängiges und starkes Format zu schaffen. PNG muss die GIF-Funktionalität wieder aufnehmen und weiter entwickeln. Derzeit ist PNG auf dem Weg zur W3C-Empfehlung und zum ISO/IEC-Standard 15984.
Funktionalität	Zusätzlich zu GIF-Funktionalitäten soll PNG unterstützen: <ul style="list-style-type: none"> - Bilder in „TrueColor“ bis zu 48 Bits/Pixel - Bilder in Grau Stufen von mehr als 16 Bits/Pixel - eine Text-Zeile (als Masken verwendete Transparenzfelder) - Informationen zu den Gamma-Werten des Bildes (= unveränderliche Anzeigewerte) - Integritäts-Check (Angaben zu fehlerhaften Stellen) - eine schnelle progressive Anzeige
Informatikvorteile	Leistungsstärker als GIF
Informatiknachteile	bis zum heutigen Tag keine bekannt
Archivvorteile	
Archivnachteile	Vgl. das zu GIF und JPEG Gesagte
Weitere Informationen	http://www.w3.org/Graphics/PNG/

Format	TIFF
	Tagged Image File Format
Status	Proprietäres Format (Adobe), de facto Standard
Typ	Grafikformat für Rastergrafik

Geschichte	Anfangs entwickelt als eine Austauschnorm für graphische Daten (hauptsächlich von digitalisierten Bildern), führte seine Universalität dazu, dass es zur Zeit das bevorzugte Format der Datenspeicherung für graphische Daten ist.
Funktionalität	TIFF definiert eine Reihe von Dateibeschreibung, die die Wiedererkennung der meisten zweidimensionalen Dateiformate erlauben. TIFF definiert Gruppentypen entsprechend den digitalisierten Daten. Persönliche „Tags“ können hinzugefügt werden, um besondere Arten von Bild zu definieren. Die große Bandbreite der Definitionen hat zu einem Normungsantrag für die professionellen graphischen Anwendungen (preprint) unter dem Namen TIFF-IT (TIFF for Image Technology) geführt, momentan in Arbeit im ISO/TC130/WG2.
Informatikvorteile	
Informatiknachteile	Hoher Speicherbedarf, insbesondere bei Farbe
Archivvorteile	De facto Standard
Archivnachteile	
Weitere Informationen	www.adobe.com/

Format	CGM
	Computer Graphics Metafile
Status	Internationale Norm (ISO 8632:1992)
Typ	Austausch- und Speicherformat für Vektorgrafik
Geschichte	
Funktionalität	CGM ist die einzige Norm, die universell für die Beschreibung graphischer Daten in Vektorform benutzt wurde. Sie definiert drei Kodierniveaus: - Clear text encoding (direkt lesbar) - Character encoding (nach ASCII-Norm) - Binary encoding (Bit mode, nicht direkt lesbar).
Informatikvorteile	
Informatiknachteile	
Archivvorteile	
Archivnachteile	
Weitere Informationen	

3. Datenbank-Formate

J.-M. Comment, Schweizerisches Bundesarchiv

Es gibt mehrere Arten von Datenbanken: hierarchisch, relational, semantisch und objektbezogen. Die Art der Datenbank wird spezifiziert nach dem Typ der logischen Datenorganisation. Momentan sind die relationalen Datenbanken am meisten verbreitet.

Dieselben Implementierungskonzepte werden auf großen Systemen (Großrechner, Hosts) wie in der Welt der Mikroinformatik angewendet. Die Unterschiede bestehen in der Anzahl der Datensätze, die eine Datenbank speichern kann (größer bei Oracle als bei Access), und in der durch eine andere physische Verwaltung der Daten erreichten Verarbeitungsgeschwindigkeit.

Die Datentypen in Tabellenprogrammen (wie Excel), basieren nicht auf den SQL-Standards, sondern nehmen Typen der Programmiersprache C wieder auf.

An der Basis benutzen alle Arten von Datenbanken den SQL-Standard (ISO 9075), was eine weitgehende Standardisierung bedeutet, aber die verschiedenen Hersteller haben ihre eigenen SQL-Dialekte entwickelt. Man muss jedoch sagen, dass die gemeinsame Basis ziemlich groß ist.

Referenz: SQL, The Standard Handbook, Stephen Cannan - Gerard Otten, McGraw-Hill Book Company, 1993, ISBN: 0-07-707664-8

Vorgestellte Arten von Daten: character, bit, exact numeric, approximate numeric, datetime, interval

Vermerk: Angesichts ihrer Komplexität wird die Archivierung von Daten aus Datenbanken Gegenstand einer spezifischen Darstellung sein müssen.

Datenart	Character
Status	Standard
Typ	Wiedergabeformat
Geschichte	E. F. Codd (1970) IBM
Funktionalität	Definiert Schriftzeichenfolgen
Informatikvorteile	Vereinheitlichung: SQL (ISO 9075)
Informatiknachteile	Dialekte: z.B. Microsoft Jet SQL-Datentyp
Archivvorteile	
Archivnachteile	
Weitere Informationen	

Datenart	Bit
Status	Standard
Typ	Wiedergabeformat
Geschichte	E. F. Codd (1970) IBM
Funktionalität	Definiert Bitketten (0 und 1)
Informatikvorteile	Vereinheitlichung: SQL (ISO 9075)

Informatiknachteile	Dialekte: z.B. Microsoft Jet SQL-Datentyp
Archivvorteile	
Archivnachteile	
Weitere Informationen	

Datenart	exact numeric
Status	Standard
Typ	Wiedergabeformat
Geschichte	E. F. Codd (1970) IBM
Funktionalität	Definiert exakte Zahlenwiedergaben
Informatikvorteile	Vereinheitlichung: SQL (ISO 9075)
Informatiknachteile	Dialekte: z.B. Microsoft Jet SQL-Datentyp
Archivvorteile	
Archivnachteile	
Weitere Informationen	

Datenart	approximate numeric
Status	Standard
Typ	Wiedergabeformat
Geschichte	E. F. Codd (1970) IBM
Funktionalität	Definiert Zahlen, die nicht immer eine exakte Wiedergabe erfahren
Informatikvorteile	Vereinheitlichung: SQL (ISO 9075)
Informatiknachteile	Dialekte: z.B. Microsoft Jet SQL-Datentyp
Archivvorteile	
Archivnachteile	
Weitere Informationen	

Datenart	Datetime
Status	Standard
Typ	Wiedergabeformat
Geschichte	E. F. Codd (1970) IBM
Funktionalität	Definiert Datumsangaben mit einem unterschiedlichen Grad an Präzision
Informatikvorteile	Vereinheitlichung: SQL (ISO 9075)

Informatiknachteile	Dialekte: z.B. Microsoft Jet SQL-Datentyp
Archivvorteile	
Archivnachteile	
Weitere Informationen	

Datenart	Interval
Status	Standard
Typ	Wiedergabeformat
Geschichte	E. F. Codd (1970) IBM
Funktionalität	Definiert Zeitintervalle
Informatikvorteile	Vereinheitlichung: SQL (ISO 9075)
Informatiknachteile	Dialekte: z.B. Microsoft Jet SQL-Datentyp
Archivvorteile	
Archivnachteile	
Weitere Informationen	

4. Audiovisuelle Formate

Jean- Jacques Eggler, Adjunkt, Stadtarchiv Lausanne

4.1 Das Digitale Video und seine Formate

Die Methoden der Übertragung analoger Signale

Ein analoges Videobild in Farbe kann durch Mischung der drei Primärfarben rot, grün und blau erstellt werden. So wie das menschliche Auge in der Praxis nicht alle Helligkeits-Informationen auf dieselbe Art und Weise wahrnimmt, kann man die folgenden Übertragungswege eines Bildes unterscheiden:

- Die RGB-Verbindung :

Die drei Primärfarben werden getrennt durch drei verschiedene Kanäle übermittelt, die jeder eine Bandbreite von 6 MHz erfordern. Beispiel: Verbindung zwischen einem PC und seinem Bildschirm.

- Die zusammensetzende Verbindung (YUV):

Um den Forderungen hinsichtlich der Erfassung und der Verbreitung zu genügen, für die die RGB-Übertragung sich nicht empfiehlt, sind die drei Basisfarben in einer anderen Form neu zusammengestellt worden, nämlich ein Leuchtdichtensignal (Y) einerseits, das das Signal schwarz und weiß transportiert, und zwei Farbsignale (U und V) für die Farbinformationen der anderen Farben. Da das menschliche Auge für schwarz und weiß empfänglicher ist als für andere Farben, können die Informationen über Farben ohne einen für das menschliche Auge sichtbaren Verlust sich auf 25% dessen beschränken, was für die Angabe der Helligkeit notwendig ist. Die für die Farben benutzte Bandbreite wird daher auf je 1,5 MHz begrenzt, während jene für die Helligkeit bei 6 MHz bleibt.

Beispiel: Magnetoskope und Betacam-Kameras. Diese Ausrüstung steht auch am Anfang der meisten digitalen Formate, wie wir weiter unten sehen werden.

Die zusammengesetzte Verbindung:

Für die Bedürfnisse des Sendens auf drahtlosem Weg war es notwendig, die Gesamtheit aller auf ein Bild bezüglichen Informationen in einem Signal zusammenzufassen. Die Methode der Kombination dieser Informationen hat drei durch das Fernsehen benutzte Standards hervorgebracht:

- PAL (Phase Alternating Line), im Jahre 1963 eingeführte deutsche Verfahrensweise;
- SECAM (SEquentiel Couleur A Mémoire), im Jahre 1967 eingeführte französische Verfahrensweise;
- NTSC (National Television System Committee oder Never Twice the Same Color), im Jahre 1953 eingeführte amerikanische Verfahrensweise.

Der Vorteil des Transports des Bildes in nur einem Signal geht, im Vergleich zu den vorher genannten Verbindungsarten, selbstverständlich auf Kosten der Bildqualität.

Beispiel: die durch VHS- oder U-Matic-Recorder registrierten Signale

- Die „verbessert“ zusammengesetzte Verbindung:

Die Bildinformationen sind hier auf Hardwareniveau getrennt in jene, die die Helligkeit betreffen, einerseits, und jene, die die Farben betreffen, andererseits. Letztere werden in

PAL oder in SECAM übermittelt. Man erhält so eine bessere Definition des Bildes als mit der klassischen zusammengesetzten Verbindung.

Beispiel: die Geräte S-VHS und Super-8.

Die Wiedergabe-Definitionen von Videobildern

Im Gegensatz zum Kino, das feste Bilder mit der Geschwindigkeit von 24 Bildern pro Sekunde benutzt, teilt das Video jedes Bild in zwei Halb-Bilder (die Raster), die sukzessiv angezeigt werden. Das Bild wird Zeile für Zeile in zwei aufeinanderfolgenden Abschnitten analysiert: ein erster Raster umfasst die Gesamtheit der geraden Zeilen und ein zweiter Raster, der sich in den ersten einordnet, umfasst die Gesamtheit der ungeraden Zeilen. Als Ergebnis erhält man 625 Zeilen (525 Zeilen in den USA) und man spricht von einem „Zeilensprung- oder Halbbildverfahren“ (Interlaced-Verfahren). Jedes vollständige Bild wird nun 25 Mal pro Sekunde angezeigt, um den Eindruck von Kontinuität in den Bewegungen zu gewährleisten.

Alle derzeitigen Fernsehsysteme sowie alle analogen und digitalen Videorecorder für Magnetbänder erstellen interlaced Bilder.

Die Systeme, die sich auf die Norm MPEG-1 stützen (siehe weiter unten), speichern die Bilder „non interlaced“, das heißt, dass sie das Bild Zeile für Zeile nur einmal analysieren. Jedes der 25 Bilder pro Sekunde wird dann zweimal angezeigt, um die 50 Lichtreize pro Sekunde zu erzeugen, die das menschliche Auge benötigt, um den Eindruck von Bewegung zu erhalten.

Für die europäischen Formate der zusammengesetzten Übertragung (PAL, SECAM), ist ein Umfang von 625 Zeilen definiert worden, die 24 und 25 Synchronisierungs- und Informationszeilen pro Raster umfassen, was zur Folge hat, dass 49 Zeilen je Bild verloren gehen. Es bleiben also tatsächlich nur 576 genutzte Zeilen.

Die Umsetzung eines analogen Videosignals in eine digitale Datei

Ein analoges Videosignal muss in eine Folge von 0 und 1 konvertiert werden, um durch einen Rechner bearbeitet und gespeichert werden zu können. Danach muss diese Folge binärer Werte in analoge Zeichen rückgewandelt werden, damit der Mensch diese Informationen verstehen kann: das ist die Aufgabe der Digital-Analog- und Analog-Digital-Umsetzer.

Um ein elektrisches Signal zu digitalisieren, benutzt man zwei Parameter:

- die Häufigkeit der Messung der elektrischen Spannung (wieviel Mal pro Sekunde wird gemessen?) ; je höher die Messhäufigkeit, desto mehr entsprechen die Messungen dem Originalsignal;
- die Bit-Menge, die benutzt wurde, um diese Messung zu kodieren; je höher die Bitzahl desto besser ist die Dynamik des digitalisierten Signals.

Im Jahre 1982 wurde die Norm CCIR 601 - besser bekannt unter dem Namen „Standard 4:2:2“ - als Norm für die Digitalisierung professioneller Videos beschlossen. Diese definiert eine Messungshäufigkeit von 13,5 MHz für die Helligkeit (Y) und von 6,75 MHz für jedes Farbsignal (U und V). Sie hat ebenfalls die Bit-Mengen, die für diese Operation benutzt werden, zunächst auf 8 Bits festgelegt, später dann erlaubt, zu 10 Bits überzugehen. So kann man die für die Aufbereitung eines vollständigen Bildes benötigte Zeit berechnen:

$(13,5+6,75+6,75) \times 10 \text{ Bits} = 270 \text{ Mbits/Sekunde}$ (216 Mbits/Sekunde, wenn man auf 8 Bits kodiert).

Diese Norm bestimmt auch, dass jede Zeile 720 Pixel für die Helligkeit enthalten muss und 360 Pixel für jede der Farbkomponenten. Das bedeutet, dass jedes zweite Pixel keine Farbinformationen enthält; es wird in schwarz und weiß übertragen. Ausgehend von diesen Grundsätzen addiert sich der Bedarf der für die Bildschirmanzeige genutzten digitalisierten Daten auf 166 Mbits/Sekunde oder, ausgedrückt in Megabyte, 20,75 MB/Sekunde.

	Pixel je Zeile		Zeile je Bild		Bild pro Sekunde		Bit pro Pixel		Bedarf in Mbits/Sekunde
Helligkeit (Y)	720	x	576	x	25	x	8	=	83
Farbe (U)	360	x	576	x	25	x	8	=	41,5
Farbe (V)	360	x	576	x	25	x	8	=	41,5
Total									166

Der Standard 4:2:2 verursacht also für eine professionelle Bildqualität ein beträchtliches Volumen digitaler Daten. Zum Vergleich: ein Rechnerbildschirm von 800 mal 600 Pixel enthält 480.000 Pixel. Um ein Bild in 16 Millionen Farben zu kodieren, muss man 24 Bit benutzen, das heißt 3 Bytes pro Pixel. Ein bildschirmfüllendes Bild „wiegt“ also 1,44 MB, multipliziert mit 25 ergibt das 36 MB für eine Videosequenz von einer Sekunde. Das ist noch mehr als der professionelle Standard 4:2:2.

Selbst wenn heute Technologien theoretisch Ausgaben digitaler Daten von 80 MB/s und sogar 160 MB/s erlauben (SCSI-Karten und SCSI-Festplatten zum Beispiel), bleibt das Problem der Übertragung dieser Volumina in der Praxis bestehen. Die einzige Lösung, diesen Ausstoß zu reduzieren, besteht auch heute noch in der Kompression der Daten (Bit Rate Reduction in Englisch).

Die Kompression der Daten

Die Norm 4:2:2 wandelt die Daten, selbst wenn sie nicht die Gesamtheit der Informationen eines Farbbildes übermitteln, trotzdem nicht um. Es handelt sich vielmehr in der Tat um ein Format zur Übermittlung nicht-komprimierter Daten. Wenn man den Datenumfang reduzieren will, könnte man zum Beispiel die Anzahl der Bilder pro Sekunde vermindern. Aber man muss wissen, dass bei weniger als 15 bis 18 Bildern pro Sekunde das menschliche Auge beginnt, Standbilder zu sehen und den Bildruck wahrnimmt. Man könnte die Anzahl Pixel je Bild reduzieren. Ebenso könnte man die Farben auf einer weniger großen Palette kodieren (64.000 anstatt 16 Millionen). Aber das ginge auf Kosten der Bildqualität.

Die Methoden zur Kompression der Daten beruhen auf der Tatsache, dass eine Videosequenz viele, von einem Bild zum anderen, redundante Informationen enthält. Man unterscheidet zwischen zwei Arten von Kompression:

- räumliche (oder intra-Bild): die Informationen sind ähnlich oder wiederholen sich in benachbarten Zonen von einem zum anderen Bild;
- zeitliche (oder inter-Bilder oder inter-Raster): die Informationen ähneln sich oder wiederholen sich in der Zeit von einem Bild zum anderen.

Mehrere Kompressionsmethoden sind ausgearbeitet worden. Eine erste Kategorie von Kompressionstechniken, nicht verlustbehaftet oder umkehrbar genannt, vermeidet es, dieselbe Information mehrmals zu übermitteln. Leider sind die hierdurch erzielten Verdichtungsverhältnisse kaum bemerkenswert. Wenn man den Datenumfang beträchtlich reduzieren will, muss man Kompressionsalgorithmen anwenden, die Informationsverluste bewirken. Man wählt gezielt die Art von Informationen aus, die verloren gehen dürfen oder verschlechtert werden können, und es trotzdem ermöglichen, Bilder von einer solchen Qualität wieder aufzubauen, dass das menschliche Auge sie nicht vom Original unterscheiden kann.

Ein solcher Algorithmus ist Discrete Cosinus Transformation (DCT), der die räumliche Darstellung eines Pixel-Blocks in eine Darstellung unterschiedlicher mathematischer Formen umwandelt. In sich selbst komprimiert diese Methode kein Bild, aber sie stellt es einfach in einer Form dar, die ihre Kompression vereinfacht.

Codec

Komprimierte Bilder wieder herzustellen erfordert eine genügend schnelle Dekompression, um einen richtigen Anzeigefluss zu gewährleisten. Der Vorgang der Kompression/Dekompression beinhaltet demgemäß zwei Schritte, Kodieren und Dechiffrieren, die zusammen einen Codec darstellen. Das Kodieren macht Gebrauch von den Kompressionsalgorithmen, die komplexe Berechnungen durchführen, entweder in Stapelverarbeitung mit sehr leistungsstarker Software und Hardware (insbesondere der Prozessor), oder in Echtzeit unter Hinzufügung einer speziellen Zusatzkarte. Die Wiederherstellung benutzt einen Algorithmus, der dem, der für die Kompression gebraucht wurde, entgegengesetzt ist. Diese Operation müsste einen Umfang von 25 Bildern pro Sekunde garantieren, was nicht immer der Fall ist.

Die Kompressionsstandards

Es wurden bereits mehrere Kompressionsstandards entwickelt. Wir geben nachfolgend eine Liste der gebräuchlichsten Kompressionsmethoden sowie ihrer Anwendungsbereiche. Diese Liste erhebt keinen Anspruch auf Vollständigkeit.

JPEG (Joint Photographic Experten Group)

Es handelt sich um einen internationalen Standard für die Kompression von Farbbildern, die in digitaler Form gespeichert wurden. Ursprünglich war dieser Standard für die Welt des Drucks und des Lichtsatzes gedacht. Er akzeptiert irgendeine vertikale oder horizontale Definition und erlaubt eine Anzahl von Bits je Pixel zwischen 1 und 32. Die benutzte Kompressionstechnik basiert auf der Discrete Cosinus Transformation (DCT). Jedoch, geplant für die Standbilder, ist er für die Kompression von Videosequenzen nicht geeignet.

Moving JPEG (oder M-JPEG)

Hierbei handelt es sich um eine Erweiterung von JPEG, die es erlaubt, Bildsequenzen zu behandeln. Sie beschränkt sich darauf, eine Videosequenz wie eine Folge von Standbildern zu betrachten, die jedes für sich getrennt mit JPEG komprimiert wurden. Das Format M-JPEG ist in Wirklichkeit nicht genormt; jeder Hersteller verschlüsselt die Bilder auf seine Weise, was die Weitergabe einer Datei von einer Maschine zur anderen erschwert.

MPEG (Moving Pictures Experts Group)

Solange man sich darauf beschränkt, Videosequenzen zu komprimieren, indem man jedes Bild einzeln betrachtet, kann der Kompressionsfaktor schwerlich 4:1 überschreiten, wenn man beabsichtigt, ein mit professionellem Gebrauch kompatibles Qualitätsniveau beizubehalten. Um höhere Kompressionsfaktoren zu erreichen, muss man auf den Ähnlichkeiten aufbauen, die zwischen mehreren aufeinanderfolgenden Bildern bestehen. Diese Feststellung hat zur Entwicklung des MPEG-Standard geführt.

Ursprünglich sah der MPEG-Standard vier Niveaus vor:

MPEG-1: bestimmt für Multimediaanwendungen;

MPEG-2: Erweiterung von Mpeg-1, die höhere Bildqualität ermöglicht;

MPEG-3: bestimmt für das hochauflösende Fernsehen;

MPEG-4: bestimmt für die mobile Kommunikation, hat nichts mit der Welt des professionellen Videos zu tun.

MPEG-1 wird auch von dem JPEG-Standard abgeleitet; benutzt seine räumliche Kompression, zu der er Algorithmen zeitlicher Kompression hinzufügt. Ziel ist es, Bilder in VHS-Qualität zu produzieren und dabei auf einen Datenumfang von 1,2 Mbits/Sekunde herunterzubrechen (1,5 Mbits/Sekunde, einschließlich Ton). Die Norm Mpeg-1 ist im

Jahre 1993 als Norm ISO 11172 angenommen worden. Die Datenträger für dieses Datenformat sind CD-ROM, Cd-I und Video-CD.

MPEG-2 war geplant, um verflochtene Bildsequenzen zu behandeln, indem man räumliche und zeitliche Kompression kombinierte. Es handelt sich hier darum, Bilder von der Qualität eines zusammengesetzten Videosystems mit einer Bitgeschwindigkeit von 4 bis 8 Mbits/Sekunde zu produzieren, oder Bilder hoher Qualität mit einem Durchsatz von 10 bis 15 Mbits/Sekunde. Er wurde 1995 als Norm ISO 13818 angenommen. Sein bevorzugter Datenträger ist heute die digitale Videodisc (DVD).

MPEG-4 ist 1999 als Norm ISO 14496 in seiner ersten Version eingeführt worden. Für eine Datenübermittlung mit sehr geringem Umfang bestimmt, ist er vor allem für das Bildtelefonat über das Web bestimmt. Eine zweite Version wurde für 2000 erwartet.

Wie man sieht, bietet MPEG einen ausgedehnten Fächer von Möglichkeiten an, erscheint flexibel und erlaubt eine gute Bildqualität. Warum sollte man es folglich nicht in einer professionellen Videofertigungskette benutzen? Dafür gibt es eine Reihe von Gründen:

MPEG gehört zu den stark asymmetrischen Systemen: der Kompressionsvorgang ist viel komplexer als der Dekompressionsvorgang. Es bedarf also einer mit Abstand höheren Berechnungskraft für die Kompression als für die Dekompression. Dies stellt kein Problem dar, wenn es sich darum handelt, Bilder zu verteilen, denn man komprimiert nur an einer Stelle (bei der Emission), dann verbreitet man die Bilder, die schließlich von den diversen Empfängern dekomprimiert werden. Im Falle einer Nachbearbeitungskette sieht das ganz anders aus: hier muss man an jedem Glied der Kette komprimieren und dekomprimieren können. Außerdem ist das MPEG-System nicht dafür konzipiert worden, bildnahe Montagen zu machen, was aber einer der Hauptgründe für Nachbearbeitung ist.

Parallel zu den Arbeiten an MPEG hat der Informatiksektor seine eigenen Lösungen entwickelt, um Videos auf die Bildschirme von PCs zu bringen. Die Anzeige- und Verarbeitungsmöglichkeiten erlaubten vom Ende der achtziger Jahre an, Standbilder hoher Qualität anzuzeigen und eine elementare Animation zu schaffen.

Quicktime (MOV-Dateien)

Das 1991 erschienene Quicktime ist ins System 7 von Macintosh integriert worden. Es handelt sich um eine Entwicklungs- und Ausführungsumwelt, die es erlaubt, klassische Dateien mit Ton- oder Videosequenzen zu verbinden.

Quicktime umfasst hauptsächlich vereinheitlichte Datenformate, Verfahrensweisen der Kompression-/Dekompression und eine spezifizierte Benutzerschnittstelle. Die benutzte Systemerweiterung basiert auf dem Grundsatz der Beibehaltung der Zeitgleichheit von Daten und führt also die Zeit als Hauptelement des Betriebssystems ein.

Quicktime ist vor allem geplant als eine sehr offene Sammlung von Spezifizierungen, die fähig ist, eine große Anzahl Entwicklungen im Bereich von Hard- und Software leicht zu integrieren, ohne die bestehenden Anwendungen wieder in Frage zu stellen.

Video für Videowindows (AVI-Dateien)

Ein durch Microsoft vermarktetes Softwarepaket, das die Speicherung und die Wiederherstellung animierter Videos auf dem PC ermöglicht. Die Ausführung von Videosequenzen ist ohne spezifische Ausstattung möglich, während die Speicherung mit Hilfe einer Digitalisierungskarte gemacht werden muss. Eine breite Kompatibilität ist gegeben dank der Definition von Spezifizierungen, die es erlauben, Komponenten für andere Funktionalitäten zu integrieren. Somit wird Video für Windows nach einer Reihe von Pilotprojekten fähig sein, Peripheriegeräte für unterschiedlichste Funktionalitäten zu un-

terstützen und dadurch soweit wie möglich die fehlenden Funktionen auf dem Softwareniveau zu ersetzen.

Ebensowenig wie Quicktime ist Video für Windows ein Kompressionsalgorithmus. Es handelt sich eher um eine standardisierte Schnittstelle zwischen der Hardware und den Kodier- und Kompressionsverfahrensweisen, die API (Programmierungsschnittstellen) anbieten, ziemlich unabhängig von der Hardware. Jedoch hat Microsoft ebenso wie Apple an verschiedenen Situationen angepasste Kompressionsalgorithmen definiert (Microsoft Video 1, Microsoft RLE compressor) und integriert auch solche, die von Dritten vorgeschlagen wurden, wie z.B. Indeo von Intel.

Mehr als eine Kompressionstechnik liefert Video für Windows in der Tat eine gemeinsame Plattform, auf der sich verschiedene Kodierverfahrensweisen artikulieren können.

Das digitale BETACAM

Vorgeschlagen durch Sony wird es von JPEG-Techniken abgeleitet, die sorgsam an die Nachbearbeitung hoher Qualität angepasst worden sind. Es benutzt Videosignale 4:2:2 und verwendet einen Faktor von 2:1 für eine schwache Kompression. Jeder Raster bietet, getrennt komprimiert, alle Montagemöglichkeiten an, die für einen professionellen Gebrauch gefordert werden. Sein Hauptnachteil ist sein schwacher Kompressionsfaktor. Es wird hauptsächlich für Anwendungen im Rundfunkbereich benutzt.

Das digitale S

Durch JVC eingeführt, benutzt dieses System die Digitalisierungsnorm 4:2:2 und die Kompressionstechnik DCT (Discrete Cosinus Transformation) vor allem für eine räumliche Kompression (Verhältnis der Kompression von 3,3:1). Es gewährleistet so einen Durchsatz von 50 Mbits/Sekunde für eine professionelle Bildqualität. Es ist kompatibel mit dem Format S-VHS.

Die digitale Videokassette (DVC)

Es handelt sich um ein neues Format für digitale Videokassetten, das durch das HD Digital VCR Consortium entwickelt wurde, das anfänglich von Sony, Matsushita, Philips, Thomson, Toshiba, Hitachi, JVC, Sanyo, Sharp und Mitsubishi gebildet wurde. Diese 10 Gesellschaften haben ihre Kräfte vereint um im Juli 1993 die erste Spezifizierung des Formates DV vorzuschlagen. Später, haben sich zahlreiche andere Gesellschaften dem Konsortium angeschlossen, das heute mehr als sechzig Mitglieder zählt. Dieses neue Format benutzt einige Elemente des JPEG-Standards für das Kodieren der Videos. Die Kompression DV behandelt nur die räumlichen Redundanzen innerhalb des vollständigen Bildes. Sie versucht nicht, die zeitliche Redundanz zu reduzieren wie es MPEG tut. Sie benutzt einen Kompressionsfaktor von 5:1. Sie erlaubt also, eine hervorragende Qualität des Bildes in der ersten Generation zu erhalten, was ideal ist für einen Gebrauch durch die breite Öffentlichkeit.

4.2 Der digitale Ton und seine Formate

Genauso wie die Bilder können Schallsignale nach demselben Grundsatz digitalisiert werden, der zwei Phasen integriert:

- das Stichprobenverfahren, das heißt die Häufigkeit der Messung des analogen Signals pro Sekunde durchgeführt in regelmäßigen Intervallen;
- die Messung dieser Messungen, das heißt die Zuteilung des Wertes an jede Messung; dieser Wert kann einer veränderlichen Anzahl von Positionen (oder Bits) zugeteilt werden, die so eine veränderliche Anzahl möglicher Werte anbietet (8 Bits, entweder 256 mögliche Werte oder 28; 16 Bits oder 65.536 mögliche Werte).

Für den Ton wird die Genauigkeit der Messung also vom Stichprobenverfahren und von der ausgewählten Messungstiefe abhängen. Je schwächer die Häufigkeit und die Tiefe, desto weniger entspricht das Ergebnis dem Originalsignal. Diese Parameter können also eine erste Quelle von Informationsverlusten sein.

Die nicht-komprimierten Formate

Audio CD

Durch Philips und Sony gemeinsam entwickelt, wird die Audio CD seit 1982 vermarktet und hat sich heute als musikalischer Tonträger als Ersatz für die Vinyl-Platte etabliert. Der digitalisierte Ton wird in einer Häufigkeit von 44,1 kHz (das heißt 44100 Mal pro Sekunde) auf 16 Bits gespeichert, was eine Datenmenge von 1,4 Mbit/Sekunde (oder 172 KB/Sekunde) bedeutet. Seine Spezifikationen die immer noch Philips gehören, sind auch bekannt unter der Benennung „Red Book“. Man muss hier hervorheben, dass das CD-Erfassungsformat nicht wie eine normale Informatikdatei behandelt werden kann (kopieren oder auf dem Rechner direkt ändern). Es bedarf einer maschinenorientierten Lesesoftware, um auf einem Rechner die musikalischen Stücke (im Englischen „tracks“ genannt) zu lesen.

WAVE (Waveform Audio File Format)

Dieses Format ist das bekannteste digitale Audio-Format in der Windows-Welt. Im Rahmen der RIFF-Spezifizierungen (Resource Interchange File Format) durch Microsoft und IBM entwickelt, benutzt es die Konvertierungsnorm PCM (Pulse Code Modulation), die durch die Internationale Union für Telekommunikation im Jahre 1990 aufgestellt worden ist. Im Gegensatz zum Audio-CD-Format kann dieser Dateityp auf einem PC mit einer speziellen Tonbearbeitungssoftware (z.B. SoundForge) kopiert oder bearbeitet werden. In der Macintosh-Welt entspricht diesem Format das proprietäre Apple AIFF-Format (Audio Interchange File Format).

Da das Leistungsvermögen des PC ständig größer wird, ist es heute möglich, ein Schallsignal in einer höheren Häufigkeit zu digitalisieren (48 kHz oder 96 KHz) und gleichzeitig eine größere Abtastrate zu benutzen (24 Bit). Aber natürlich hat diese Qualität einen Preis: die Dateien, die sich daraus ergeben, haben einen sehr großen Umfang.

DAT (Digital Audio Tape)

Eingeführt durch Sony im Jahre 1987 ist der DAT-Standard ein digitales Erfassungsformat das im normalen Betrieb eine Abtastrate von 48 kHz und eine 16-Bit-Kodierung nutzt, also eine Qualität erzeugt, die höher ist als jene von CD-Audio. Die Erfassung erfolgt auf einem Magnetband über einen rotierenden Kopf (daher die anfängliche Benennung "R-DAT" für Rotary-Digital Audio Tape), was eine hohe Informationsdichte ohne Fehlerrisiko erlaubt. Die DAT-Kassette ist kleiner als eine klassische Audio-Kassette und kann eine Hördauer von bis zu 3 Stunden bieten.

Die komprimierten Formate

MP3 (MPEG 2.5 Audio Layer III)

Alle Geräusche, die uns umgeben, setzen sich aus einer Mischung von Tausenden unterschiedlichen Frequenzen zusammen (je höher die Frequenz desto spitzer/schriller der Ton). Das Format MP3 analysiert die Geräusche, um die Frequenzen zu finden, aus denen sie sich zusammensetzen, und beschränkt sich dann darauf, einen kleinen Teil dieser Frequenzen zu speichern. Dadurch erreicht man eine Komprimierung bis zum Faktor 12 bei einer CD-Audio nahen Abhörqualität. MP3 wurde 1995 als Norm ISO 13818:1995 von ISO/IEC gebilligt.

Je nach der gewünschten Verbreitungsqualität kann man mehrere Arten einer MP3-Datei produzieren, die eine mehr oder weniger große Datenmenge erzeugen (von 8 bis

320 kbit/s). Die beste Qualität einer MP3-Datei entspricht dem von CD-Audio, was den Erfolg von MP3 erklärt. Allerdings stellt MP3 auch ein ernstes rechtliches Problem in Bezug auf musikalische Raubkopien dar.

Real Audio (RM Dateien)

Eigentum der amerikanischen Firma RealNetworks. Dieses Format ist im Jahre 1995 erschienen. Es beruht auf den Grundsätzen von ADPCM (Adaptive Differential Pulse Code Modulation), die es verbessert. Anstatt alle Werte nach einer gewissen Häufigkeit zu messen, beschränkt man sich darauf, nur die Unterschiede zwischen diesen Werten zu messen. Anfänglich war die Tonqualität nicht sehr gut, doch wurde dieses Format bis heute besonders gut an eine Übertragung über das Internet angepasst, das heißt an einen sehr geringen Platzbedarf. Die Ton-Datei wird nicht gänzlich vom Web-Host heruntergeladen, sondern in kleinen Paketen je nach Fortschritt des Abspiels. Dies ist besonders vorteilhaft für umfangreiche Musikstücke oder um Radioprogramme direkt zu hören. Dieses Format unterscheidet sich gerade in diesem Punkt von seinen Konkurrenten, die zu einem vollständigen Herunterladen der Datei zwingen, bevor sie abgespielt werden können.

Genau wie das Format MP3 erlaubt auch Real Audio, mehrere Versionen ein und derselben Tondatei je nach der gewünschten Abhörqualität zu erzeugen.

Minidisc

Herausgebracht durch Sony im Jahre 1992. Minidisc ist sowohl ein Erfassungsformat als auch ein besonderer wiederbeschreibbarer magneto-optischer Datenträger mit einem Durchmesser von 64 mm. Die Kapazitätsgrenze liegt bei 80 Min. Die benutzte Stichprobenhäufigkeit liegt bei 44,1 kHz, und die Messung wird auf 16 Bit durchgeführt. Die Kompression der Informationen, ein Sony eigenes Verfahren, gehört zum Typ ATRAC (Adaptive Transform Acoustic Coding); es benutzt den Maskierungseffekt und die eingeschränkte Hörfähigkeit des menschlichen Ohrs. Nach Analyse der Komponenten der Häufigkeit des Signals identifiziert das ATRAC-System jene, die unterhalb der Hörschwelle liegen, und eliminiert sie. Ebenso hört das menschliche Ohr von zwei Tönen, die gleichzeitig gesendet werden, einer sehr stark und der andere schwächer, nur den stärkeren Ton. Somit kann, ab einer gewissen Schwelle, der schwächste Ton unterdrückt werden.

Das Abspielsystem ist ebenfalls Sony eigen. Seine Besonderheit liegt in der Benutzung eines stoßfesten Speichers, der es ermöglicht, die Wirkungen von Vibrationen oder Stößen zu minimieren.

Bibliographie

Multimediatdaten allgemein:

<http://www.cndp.fr/notestech/25/donmulti.htm>

Zu digitalen Video-Formaten:

http://www.cndp.fr/notestech/27/nt027_0.htm

Zu digitalen Ton-Formaten:

<http://perso.club-internet.fr/farzeno/dat.htm>

Zu MPEG-Formaten:

www.mpeg.org/MPEG/index.html

Digitale Video-Standards

Format	Hersteller	Trägergröße	Quantifizierung	Verhältnis	Bedarf	Codierung	Datenträger	Transport
D1	Sony/BTS	19 mm	4:2:2/8-10bit	1:1	270Mbits/s	ITU-R 601	Metall	SDI
D2	Sony	19 mm	Zusammengesetzt digital	1:1	-	-	Metall	zusammengesetzt digital I/O
D3	Panasonic	1/2-Zoll	zusammengesetzt digital	1:1	-	-	Metall	zusammengesetzt digital I/O
D5	Panasonic	1/2-Zoll	4:2:2	1:1	-	-	Metall	SDI
D6	-	19 mm	-	-	-	-	-	-
D7(D VCP RO50)	Panasonic	6.35 mm	4:2:2	3.3:1	50Mbits/s	DoubleDV	Metallteilchen	SDI
DCT	Ampex	19 mm	4:2:2	2:1	-	DCT	Metall	SDI
Digital Beta-cam	Sony/Thomson	1/2-Zoll	4:2:2	2.2:1	-	DCT	Metall	SDI
Beta-cam SX	Sony	1/2-Zoll	4:2:2 PGML	10:1	18Mbits/s	MPEG-2 Gop : IP	-	QSDI/SDI
DVC AM	Sony	6.35 mm	4:2:0	5:1 /	25Mbits/s	DV 1 frame	Metall	SDI/QSDI/1394
DVC PRO	Panasonic	6.35 mm	4:1:1	5:1	25Mbits/s	DV 1 frame	Metallteilchen	SDI
DVC PRO50	Panasonic	6.35 mm	4:2:2	3.3:1	50Mbits/s	Double-DV 1 frame	Metallteilchen	SDI
Digital-S	JVC	1/2-Zoll	4:2:2	3.3:1	50Mbits/s	Double-DV	-	SDI
DV	consortium	6.35 mm	4:2:0	5:1	25Mbits/s	DV 1 frame	Metall	S-Video/1394

Format	Hersteller	Trägergröße	<i>Quantifizierung</i>	<i>Verhältnis</i>	<i>Bedarf</i>	Codierung	Datenträger	Transport
D-VHS	consortium	1/2-Zoll	-	-	-	MPEG-2	-	-
D-VHS HD	JVC	1/2-Zoll	-	-	-	720p/1080i	-	-
W-VHS	JVC	1/2-Zoll	-	-	-	SMPTE240M	-	-
Media100	DataTranslation	-	-	2:1	-	M-JPEG	HD	YUV
AVR (77)	Avid	-	-	2:1	-	M-JPEG	HD	YUV
DVD	consortium	12cm	4:2:0 PGML	variable	5 Mbits/s environ	MPEG-2 long GOP	Polykarbonat	S-Video/1394*
VideoCD 2.0	consortium	12cm	4:1:1	variable N:12 M:3	1.5Mbits/s	MPEG-1	Polykarbonat	zusammengesetzt
Laser disc	consortium	12-Zoll	Zusammengesetzt; 1 Zoll	-	-	NTSC/PAL	PMMA	zusammengesetzt

Quelle : Jean-Charles Fouché,

URL : <http://www.mireade.com/repaiрев/savoir/essentiel/05.htm> (nicht mehr aktiv)

5. Elektronische Post (E-Mail)

von Karl-Ernst Lupprian, Generaldirektion der Staatlichen Archive Bayerns

Die elektronische Post ist in kurzer Zeit in fast allen Lebensbereichen zu einem alltäglichen Werkzeug geworden. Zwar werden wichtige E-Mails bislang noch meistens ausgedruckt und zu den Akten genommen, doch wird sich dies mit der Einführung von Dokumentenmanagement-Systemen bald ändern. Archive werden dann mit großen Mengen von E-Mail konfrontiert, die u.U. auch in elektronischer Form aufbewahrt werden müssen, wenn nämlich Papier-surrogate rechtlich nicht anerkannt werden.

Die Erkenntnis, dass es sich bei E-Mails um simple Nur-Text-Dateien handelt, ist zwar richtig, hilft aber wenig bei der Lösung des Problems ihrer Archivierung.. Hierfür müssen die Mechanismen etwas näher betrachtet werden, die für die Erzeugung, den Transport und die Bereitstellung elektronischer Post verwendet werden.

Eine E-Mail wird vom Rechner des Absenders an einen Mailserver versandt, der die Mail über das Internet an den Mailserver weiterreicht, an dem der Empfänger angemeldet ist. Diesen Transport erledigt ein Protokoll namens **SMTP** (*Simple Mail Transfer Protocol*, definiert in den RFCs³ 821, 1425, 1651, 1830, 1845, 1854, 1869, 1870 und 2197). Der Rechner des Empfängers fragt entweder den Mailserver nach vorliegender Post oder wird von ihm benachrichtigt. Für diesen Teil des Transportswegs ist heute vorwiegend das Protokoll **POP3** (*Post Office Protocol 3*, definiert in den RFCs 1081, 1082, 1225, 1460, 1725 und 1939) zuständig. Wenn die Mail beim Rechner des Empfängers angekommen ist, wird sie auf dem Mailserver gelöscht. Neuere Mechanismen sind **IMAP** (*Internet Message Access Protocol*) und **ACAP** (*Application Configuration Access Protocol*, noch in der Versuchsphase). Diese erlauben dem Mailempfänger die Verwaltung seiner eingehenden Mails bereits auf dem Mailserver, und zwar in einer weitgehend automatisierbaren Form.

Betrachtet man den Aufbau einer E-Mail, so sind grundsätzlich drei Teile zu unterscheiden: Der Kopf (Header), die Nachricht (Body) und der Umschlag (Envelope). Letzterer spielt nur für den Mailtransport eine Rolle und muss hier nicht weiter betrachtet werden.

Der Kopf enthält sämtliche Adress- und Verbindungsdaten und erlaubt damit (sofern keine Fälschung vorliegt) die Rückverfolgung des Weges, den die E-Mail genommen hat. Ferner beschreibt der Kopf die Kodierung der Nachricht (MIME-Typ, s.u.) und ggfs. Art und Position der Anhänge.

Die E-Mail-Nachricht ist eine Folge von ASCII-Zeichen, die in 7 Bits kodiert sind. Das bedeutet, dass nationale Sonderzeichen (z.B. deutsche Umlaute und das ß) anderweitig dargestellt werden müssen. Anhänge (attachments) wie z.B. Worddokumente, Bilder oder Programmdateien sind integrierte Bestandteile des Body und müssen besonders kodiert werden, um sie dem 7-Bit-ASCII-Schema anzupassen.

Ursprünglich (RFC 822) konnte man mittels E-Mail nur Textnachrichten übertragen, und diese durften nur Zeichen enthalten, die der US-ASCII-Zeichensatz erlaubt. Das änderte sich erst mit der Einführung von **MIME** (*Multipurpose Internet Mail Extensions*, RFC 2045-2049). Seitdem können auch Zeichensätze mit nationalen Sonderzeichen oder nichtlateinischen Schriften, wie sie z.B. in der Zeichensatzfamilie ISO-8859 standardisiert sind, verwendet werden. Probleme entstehen jedoch, wenn der Rechner des Empfängers einer E-Mail den vom Absender verwendeten Zeichensatz nicht oder nicht korrekt interpretieren kann. Seit MIME kann eine E-Mail-Nachricht auch mit HTML kodiert werden.

Die Anhänge werden mit dem Verfahren **Base64** in 7-Bit-ASCII-Darstellung umkodiert und beim Mailempfänger mit dem gleichen Verfahren dekodiert. Das Verfahren ist unter:

³ RFCs (Requests for Comment) enthalten verbindliche Definitionen und Festlegungen für den technischen Betrieb des Internet (vgl. <http://www.rfc-editor.org/>).

<http://home.t-online.de/home/arndt.bruenner/mathe/scripts/base64.htm> in leicht verständlicher Form dargestellt.

Bei der Archivierung einer E-Mail ist diese vollständig (also mit dem Kopf) als Nur-Text-Datei abzuspeichern. Probleme bereiten jedoch die Anhänge, da zu ihrer Dekodierung aus der Base64-Darstellung das Programm benötigt wird, mit dem sie ursprünglich erstellt wurden (also z.B. Microsoft Word). Daher wird man die Anhänge zeitnah (also möglichst bald nach dem Eingang der Mail) in ein langzeitfähiges Format (z.B. TIFF oder PDF) umwandeln müssen. Diese Dateien müssen dauerhaft mit der Mail verbunden werden. Ein beachtenswerter Lösungsweg ist in den Niederlanden besprochen worden: *From digital volatility to digital permanence. Preserving email*, Den Haag 2003 (als PDF erhältlich über: <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/volatility-permanence-email-en.pdf>).