

## **Ulrike Gutzmann, Ulrich Kamp, Christian Keitel, Antje Scheiding für den Arbeitskreis "Elektronische Archivierung" in der Vereinigung deutscher Wirtschaftsarchivare e.V. (VdW) (1)**

### **Praktische Lösungsansätze zur Archivierung digitaler Unterlagen: "Langzeitarchivierung" und dauerhafte Sicherung der digitalen Überlieferung**

#### ***Einleitung: Datenhaltung und Archivieren***

Die mit dem Ausdruck "Langzeitarchivierung" umschriebenen Bemühungen zur Erhaltung der Zugänglichkeit und Sicherheit digitaler Daten über einen langen Zeitraum nehmen die Archivare den Kollegen aus der IT nur ungern und widerstrebend ab. So stößt man in den Computerwelten allenthalben auf Archive, und ein namhaftes E-Mail-Programm fragt seine Nutzer zuweilen, ob sie "autoarchivieren" möchten. Unter IT-Spezialisten und leider inzwischen auch anderweitig wird für die elektronische Datenhaltung gerne das Wort "Langzeitarchivierung" verwendet, das allerdings häufig einen Zeitraum von maximal 10 Jahren umfasst. Von jeher haben die Archivare bei ihrer Tätigkeit die Ewigkeit im Blick – und darum für derlei kurzfristig angelegten und vordergründigen Umgang mit dem Wort "Archivierung" wenn überhaupt nur ein wissendes Lächeln übrig. Und genau das ist das Problem. Denn während die IT die Welt in immer schnelleren Zyklen mit digitalen Innovationen beglückt und vollmundig verspricht, das Archivieren – gemeint ist damit in der Regel die Datenhaltung – gleich mit zu übernehmen, verzweifeln die Fachleute in den Archiven bei dem Versuch, die dabei geborenen Daten über einen langen Zeitraum hinweg für die Nachwelt festzuhalten, und zwar lesbar, nachvollziehbar und authentisch. Dabei stehen sie zunächst vor der Frage, nach welchem Konzept eine Archivierung vorgenommen werden kann. Die Wahl geeigneter Formate für Text-, Bild-, und Tondaten ist ebenso von Bedeutung wie die Festlegung der dabei erforderlichen Metadaten. Standards finden weder für Formate noch für Metadaten Anwendung, und so findet der Archivar auf der Suche nach Anhaltspunkten kaum Hilfe. Dieser Zustand ist lange beklagt und die daraus erwachsenden Gefahren sind ausreichend beschrieben worden. Grundsätzlich kann nur in einem gemeinsamen Bemühen von IT und Archiv die langfristige bis dauerhafte Sicherung digitaler Unterlagen gelingen. Aufgabe der Archivare und Archivarinnen ist es, ihre Anforderungen klar zu formulieren und gegenüber den Mitarbeitern der IT in den Rechenzentren und Fachabteilungen zu vertreten, damit diese für die geeignete Datenhaltung sorgen. Dieses Papier bietet dafür eine Orientierungshilfe und Lösungsansätze. (2)

Der Arbeitskreis "Elektronische Archivierung" (AKEA) der Vereinigung deutscher Wirtschaftsarchivare (VdW) hat in fünf Sitzungen seit November 2004 nach Lösungsansätzen für einige grundsätzliche Probleme beim Umgang mit der digitalen Überlieferung gesucht. Ausgehend vom vielfach geäußerten Wunsch, Hinweise für die vorgesetzten Stellen zu erhalten, die den Archivarinnen und Archivaren die Türen öffnen, wenn es um die Beteiligung an Entscheidungen über neue Systeme im Unternehmen geht, bis hin zu ganz praktischen Fragen nach geeigneten Formaten und Metadaten bei der Übernahme von digitalen Unterlagen ins Archiv, wurden von den Mitgliedern des AKEA die folgenden Empfehlungen formuliert.

Dabei wurde zunächst auf der Basis der aktuellen Fachdiskussion Einigkeit darüber erzielt, dass die Empfehlungen des AKEA im Hinblick auf die Langzeitspeicherung elektronischer Unterlagen den Ansatz der Migration verfolgen. Weder die Emulation, also die Simulation einer historischen Hard- und Softwareumgebung in einem modernen System, noch die Vorhaltung historischer Hard- und Software sind Erfolg versprechend zur dauerhaften Aufbewahrung digitaler Dokumente.

Wie die Mitglieder aus ganz unterschiedlichen Archiven kommen – Unternehmens- und Wirtschaftsarchive ebenso wie kommunale und staatliche Archive, so sind auch die Empfehlungen für die Archivpraxis ganz allgemein gedacht und sollen Anwendung nicht nur in Unternehmens- und Wirtschaftsarchiven finden. Allen, die am Zustandekommen mitwirkten, ob als Teilnehmer an den Arbeitskreissitzungen oder als Referent, sei an dieser Stelle herzlich gedankt.

Der folgende Artikel fasst die Ergebnisse zusammen und gliedert sich in vier Abschnitte:

Wie sag ich`s meinen Kollegen? Begründung für die Beteiligung des Archivs bei der Einführung neuer Systeme

Was tun, wenn digitale Daten angeboten und/oder migriert werden müssen? Metadatenkatalog für die Übernahme und Erhaltung von digitalen Daten im Archiv

Auf dem Weg zum "Traumformat": Kriterienkatalog zur Bestimmung geeigneter Archivierungsformate. Welche Anforderungen müssen Formate zur langfristigen Speicherung digitaler Unterlagen erfüllen?

Welche Formate eignen sich für die Speicherung digitaler Unterlagen? Formatematrix zur Übersicht und Klassifizierung geeigneter Archivformate für Text- Bild- und Audiodokumente.

## **1. Begründung für die Beteiligung des Archivs bei der Einführung neuer Systeme**

Das wichtigste Argument für die Beteiligung des Archivs im Systementwicklungsprozess ist seine Kernkompetenz, Daten und Informationen zu strukturieren. Damit erzielt das Unternehmen Rechtssicherung und erreicht eine Kostenreduzierung, indem die aufzubewahrenden Daten auf das Wesentliche reduziert, Systembrüche in Prozessketten vermieden und Folgekosten durch die präventive Einbindung des Archivs reduziert werden.

Nur bei Berücksichtigung im Prozess der Systementwicklung kann das Archiv seinen Dienstleistungsauftrag erfüllen, die für das Unternehmen wichtigen Daten dauerhaft zu sichern, d.h. zugänglich und lesbar zu erhalten.

Aus archivischer Sicht ergeben sich dabei Anforderungen an die Systementwicklung. Voraussetzungen sind die Schaffung von Schnittstellen zur Übernahme und Archivierung digitaler Unterlagen sowie die Bewertung der Prozesse im Vorfeld durch das Archiv. Dabei muss geklärt werden, ob in einem System potentiell archivwürdige Unterlagen entstehen oder solche, die nach Ablauf der Aufbewahrungsfristen gelöscht werden können. Im Hinblick auf die dauerhafte Sicherung digitaler Daten muss im Unternehmen eine Strategie entwickelt werden, die idealerweise ein Migrationskonzept verfolgt. Migration ist erforderlich bei Änderungen von Hard- und Software beim Datenanbieter, beim Archiv oder bei der Datenstruktur. Bei der Einführung neuer EDV-Systeme ist das Archiv zu beteiligen.

Schließlich muss der Übergang von Daten in das Archiv geklärt werden, wobei Zeitpunkt, Modus und rechtliche Rahmenbedingungen geregelt werden müssen. Im Hinblick auf Kosten für das Unternehmen ist es unbedingt erforderlich zu klären, wer die Kosten für Entwicklung, Einrichtung und Betrieb der Schnittstelle und für die Datenhaltung der zu archivierenden Unterlagen übernimmt.

Regelmäßige Qualitätskontrollen gewährleisten eine hohe Sicherheit vor Datenverlusten.

## **2. Metadatenkatalog für die Übernahme und Erhaltung von digitalen Daten im Archiv**

Ist der erste Schritt erfolgreich getan, so werden dem Archiv bald auch digitale Dokumente angeboten. Außerdem gehen viele Archive dazu über, analog vorhandene Unterlagen einzuscannen, weil sie so durch Mitarbeiter und Nutzer leicht am Rechner zu lesen und auszuwerten sind, und weil das analoge Original dadurch in seiner Erhaltung geschont wird. Für die digital geborenen, wie für die digitalisierten Daten muss über die mit ihnen geführten Metadaten entschieden werden. Der gewissenhafte Umgang mit den Metadaten ist für Integrität, Authentizität und Identität aller digitalen Unterlagen unerlässlich und Voraussetzung für ihre mögliche Gerichtsverwertbarkeit.

Der folgende Katalog ist das Ergebnis der Diskussion im AKEA. Priorität hatte die Anwendbarkeit in der Praxis, die eine Reduzierung der Metadaten auf das absolute Minimum geboten erscheinen ließ. Ist der Katalog zu stark reduziert, leiden langfristig Aussagefähigkeit und Datensicherheit, ist er zu umfangreich besteht die Gefahr, dass er nicht angemessen gepflegt werden kann.

## **Metadatenkatalog für die Übernahme und Erhaltung von digitalen Daten im Archiv**

### **Präambel**

Der vorliegende Metadatenkatalog listet die Metadaten auf, die ein Archiv zur langfristigen Erhaltung digitaler Unterlagen benötigt.

Je frühzeitiger und umfassender die Übernahme der Metadaten an das Archiv geregelt wird, desto wirtschaftlicher kann sie vorgenommen werden. Eine Erhebung der Metadaten vor der Abgabe ans Archiv ist daher zu empfehlen. Um den Aufwand möglichst gering zu halten, erfolgen Erhebung und Übernahme der Metadaten wenn möglich automatisch.

*Metadaten* sind "Daten über Daten", die zum Verwalten, Verarbeiten, Sichern, Recherchieren und Nutzen von digitalen Dokumenten und ihrem Kontext benötigt werden. Ihr Inhalt ist zum Zeitpunkt der Übernahme ins Archiv nicht fest mit dem Datenträger verbunden. Es ist unerheblich, ob diese Dokumente weiterhin digital vorgehalten werden. Eine Speicherung auf analogen Speichermedien ist ebenfalls möglich.

Der vorliegende Metadatenkatalog ermöglicht eine Beschreibung von Dokumenten und Dokumentenarten unter Verzicht auf Spezialfälle. Zweck ist die Erfassung der Metadaten, die im Archiv gebraucht werden, um die Brauchbarkeit digitaler Daten im Hinblick auf ihre Nutzbarkeit, Vollständigkeit und Authentizität dauerhaft sicher zu stellen.

Die Zusammenstellung fasst die Minimalanforderungen zusammen. Eine Priorisierung und gegebenenfalls Erweiterung erfolgt durch die Archive selbst und nach deren Bedürfnissen.

Der Katalog ist eine Grundlage für die Ausgestaltung der jeweiligen innerarchivischen Organisation und für die Zusammenarbeit mit den Akten führenden Stellen und dem Vorarchiv.

Für die Authentizität der elektronischen Daten bei der Übergabe in das Archiv ist die abgebende Stelle zuständig. Eine Dokumentation über die technische und inhaltliche Entwicklung der Daten nach Übernahme ins Archiv ist notwendig.

Metadaten entstehen für drei Bereiche: Die *beschreibenden Metadaten* sind auf den Inhalt bezogen, die *erhaltenden Metadaten* auf die technische Umgebung und die *verwaltenden Metadaten* dienen zu ihrer Verwaltung im Archiv.

### **Metadatenkatalog**

#### ***Beschreibende Metadaten:***

Bestand (Ordnungseinheit im Archiv)

Titel (Kurzbezeichnung eines Dokuments)

Laufzeit / Entstehungsdatum (erstes und letztes Bearbeitungsdatum bzw. Datum der Erstellung)

Provenienz (Entstehungszusammenhang, aktenbildende Stelle)

Abgebende Stelle (Organisationseinheit, die das Dokument dem Archiv übergeben hat)

Aktenplan (Systematik der aktenbildenden Stelle)

Klassifikation (Schema der Struktur der aktenbildenden Stelle für die Gliederung von Findbüchern)

Inhalt (genauere Beschreibung des Dokuments mit Enthält- und Darinvermerk)

Stichwort / Schlagwort / Suchbegriff / Thesaurus

Archivalientyp (z.B. Akte, Foto, Film, Zeichnung)

#### ***Erhaltende Metadaten:***

Format / Version (Datenformat, Dateiformat, Träger; Protokollierung durch das System)

Datum (letztes Speicherdatum, Übernahmedatum, Erstellungsdatum)

History (Protokollierung von Veränderungen vor und nach Übernahme z.B. Migration, Verantwortlichkeiten, verwendete Hard- und Software)

Größe (z.B. in KB, MB)

Umfang (Anzahl der Dateien)

Struktur (z.B. Multimediadatei, Webauftritte, verschiedene Formate, verschiedene Bestandteile, funktionale Zusammenhänge)

Prüfsumme (Hashwert, elektronische Signatur)

Formatspezifische Attribute (Framerate, Code, z.B. Bildrate bei digitalen Bildern, Pixel, Streamingdaten, Kompressionsverfahren)

#### ***Verwaltende Metadaten:***

Rechte (z.B. Persönlichkeit / Urheber / Verwertung)

Sperrfrist (Datum, bis zu welchem ein Dokument für die öffentliche Nutzung gesperrt ist)

ID-Nummer (eindeutige vom System generierte Identifizierungsnummer)

### **3. Auf dem Weg zum Traumformat: Kriterienkatalog zur Bestimmung geeigneter Archivierungsformate**

Immer wieder ertönt aus den Archiven der Ruf nach einer Zusammenstellung der geeigneten Formate zur Speicherung von Text, Bild, Audio- und Videodateien. Um es gleich vorweg zu sagen, eine einfache Antwort auf diese Frage gibt es nicht, stets sind die spezifischen Voraussetzungen und Anforderungen des jeweiligen Archivs zu berücksichtigen. Darum muss in jedem Archiv vor der Entscheidung über die Wahl der Formate festgelegt werden, welche Kriterien der Wahl zugrunde liegen.

Der AKEA hat zu diesem Zweck einen Kriterienkatalog erarbeitet, der die unbedingt bedenkenswerten Kriterien auflistet und den Archiven Hilfestellung für die eigene Auswahl bietet. Dabei flossen insbesondere die Kriterien der Library of Congress ein, die im Internet unter [www.digitalpreservation.gov/formats/sustain/sustain.shtml](http://www.digitalpreservation.gov/formats/sustain/sustain.shtml) einsehbar sind. Zuweilen kann es vorkommen, dass sich einzelne Anforderungen widersprechen. So wird bei den meisten Textdokumenten die Forderung nach Lesbarkeit mit Basiswerkzeugen

wie einfachen Texteditoren nicht vereinbar sein mit der Anforderung nach Layouttreue, die meist nur durch die Programme erreicht wird, mit denen das Dokument erstellt wurde. Hier zeigt sich, dass die Bedeutung, die ein Archiv einer bestimmten Anforderung zuweist, für die Entscheidung über ein Archivzielformat ausschlaggebend sein muss.

Zunächst werden die verschiedenen Kriterien vorgestellt und erläutert.

In einem nächsten Schritt werden die verschiedenen Formate für Text-, Bild- und Tondokumente auf ihre Tauglichkeit im Hinblick auf die einzelnen Kriterien untersucht, bewertet und in einer Matrix zusammengestellt, die den Archiven eine Orientierungshilfe bei der Suche nach geeigneten Zielformaten für die Archivierung elektronischer Unterlagen sein soll.

Im Hinblick auf die Dateiformate für Bilder ist zu beachten, dass nur zweidimensionale Bilder und vor allem Rasterabbildungen berücksichtigt wurden, Vektorabbildungen wurden bis auf SVG außen vor gelassen. Außerdem fanden PDF und PDF/A aufgrund ihrer weiten Verbreitung Berücksichtigung. Wichtig ist ferner, dass dasselbe Format im Hinblick auf unterschiedliche Daten (Text- oder Bild) sehr unterschiedlich bewertet sein kann.

Filmformate konnten nicht berücksichtigt werden, darum sei hier lediglich der Hinweis auf das gegenwärtig als quasi-Standard anzusehende Format Digi Beta gegeben, das im professionellen Umfeld, etwa bei Rundfunkarchiven, weit verbreitet ist und zur Zeit die besten Möglichkeiten zur Archivierung digitaler Filmdokumente bietet.

Bei der Suche nach geeigneten Audioformaten muss vorab geklärt werden, ob Sprache oder Musik archiviert werden soll. Zusätzlich zum Archivformat, das bei den vorliegenden Bewertungen im Fokus stand, ist möglicherweise auch für das so genannte Vorhören ein geeignetes Format einzusetzen. In datenreduzierten Formaten, in denen Audiodaten gekapselt vorliegen, ist bei geringer Speicherkapazität für Nutzungszwecke ein erstes Anhören von Audioquellen möglich, bei dem es lediglich auf das Wiedererkennen oder "Reinhören" ankommt. Aufgrund der großen Datenmengen im Audiobereich ist hier auch die so genannte Streamingfähigkeit eines Formats nicht unwichtig. Dabei erfolgt der Zugriff auf das File über einen Streamingserver, ohne dass das gesamte File über einen möglicherweise sehr langwierigen Download geladen werden muss. Beim Streamingserver ist jeder beliebige Punkt im Quellfile ansteuerbar.

Die Forderung, für digitale Dokumente langfristig Rechtsverbindlichkeit bzw. Revisionssicherheit zu gewährleisten, wird in den Archiven, Rechtsabteilungen, aber auch von IT-Seite immer lauter. Die Nestor-Arbeitsgruppe "Vertrauenswürdige Archive – Zertifizierung", die sich sehr eingehend mit diesem Themenkomplex beschäftigte, hat im Juni 2006 einen ersten Kriterienkatalog zur Diskussion vorgelegt, der unter [http://nestor.cms.hu-berlin.de/moinwiki/AG\\_Vertrauensw%C3%BCrdige\\_Archive -  
\\_Zertifizierung](http://nestor.cms.hu-berlin.de/moinwiki/AG_Vertrauensw%C3%BCrdige_Archive_-_Zertifizierung) im Netz auffindbar ist. Auch in den Diskussionen des VdW-Arbeitskreises wurde Einigkeit darüber erzielt, dass es letztlich das Archivierungskonzept insgesamt ist, das über die Rechtsverbindlichkeit bzw. Revisionssicherheit von Informationen bzw. Dokumenten entscheidet. Dabei kommt es vor allem darauf an, zwei Bedingungen zu erfüllen, die grundlegend für eine dauerhafte Rechtssicherheit elektronischer Daten sind: Voraussetzung ist zum einen die Erfüllung der verschiedenen, unten genannten Anforderungen und zum anderen der Archivierungsprozess in den Archiven selber. So muss beispielsweise zunächst die Vollständigkeit der Informationserhaltung durch ein Dateiformat sichergestellt sein, ein Aspekt, der bereits im Kriterium Layouterhaltung vollständig enthalten ist.

Die zweite wesentliche Anforderung für die Rechtsverbindlichkeit von Dokumenten ist die Forderung, dass es Mechanismen zum Nachweis der Unveränderlichkeit von Unterlagen geben muss (Integritätsnachweis). Gerade hier erscheint aber die Auswahl eines Dateiformates als ungeeigneter Ansatz, denn sobald ein Format vollständig technisch lesbar ist, kann auch jederzeit eine Veränderung vorgenommen werden und das geänderte Dokument als vermeintliches Original wieder abgelegt werden. Daher ist die

Anforderungen des Integritätsnachweises grundsätzlich nicht durch ein Dateiformat selbst zu gewährleisten, sondern wird durch das Archivierungskonzept erreicht und muss durch Ausgestaltung des Archivierungsworkflow und durch die Bedingungen bei der Datenhaltung abgebildet werden.

Das federführend an der Physikalisch Technischen Bundesanstalt (PTB) in Braunschweig entwickelte Konzept "ArchiSafe (Langzeitarchivierung)" verfolgt ebenfalls diesen Ansatz und knüpft damit an die Ergebnisse des mit öffentlichen Geldern des Bundesministeriums für Wirtschaft und Arbeit im Jahre 2003 geförderten Projektes "ArchiSig" an, in dem wesentliche Grundlagen der rechtssicheren elektronischen Archivierung erarbeitet wurden. "ArchiSafe (Langzeitarchivierung)" definiert die Grundlagen für eine kostengünstige und skalierbare elektronische Archivlösung. Genauere Informationen dazu finden sich unter [www.archisafe.de](http://www.archisafe.de).

#### **Die Kriterien zur Auswahl von Formaten:**

##### **1. Lesbarkeit / Wiedergabefähigkeit (ohne Spezialwerkzeuge):**

Die digitale Darstellung ist offen für die unmittelbare Analyse mit Basiswerkzeugen (wie z.B. Text- oder Bildeditor oder Audioplayer, ohne Nutzung von Spezialwerkzeugen).

Textinformationen sind in Standardzeichensätzen (ASCII, Unicode) kodiert.

##### **2. Wiedergabequalität (z.B. Layouterhaltung, Farbtiefe, Klangqualität)**

Die layoutgerechte Anzeige der vollständigen Bild- und Textinformationen mit geeigneten Werkzeugen ist einfach möglich.

Layout, Struktur und Navigation im Dokument bleiben erhalten, ebenso Schrift-Definitionen und erweiterte DTP-Eigenschaften.

Die Erhaltung der Anzeige von grafischen Elementen (mathematische oder chemische Formeln, Abbildungen, Audioqualität etc.) ist sichergestellt.

##### **3. Freie Verarbeitbarkeit / Lesbarkeit durch Maschinen:**

Formate, die durch kryptografische Verschlüsselung an bestimmte Lizenzinhaber, durch bestimmte technische Schutzmechanismen (z.B. Dongle) oder für bestimmte Zeiträume eingeschränkt werden können, eignen sich nicht für die Archivierung.

Die Möglichkeit, den Inhalt von Dateien zu nutzen oder auch an neue IT-Gegebenheiten anzupassen und ggf. zu konvertieren oder an eine andere technische Bandbreite anzupassen, ist jederzeit gegeben.

Text- und binäre Byteinformationen sind in normierten Daten- bzw. Zeichensätzen (ASCII, Unicode) kodiert.

Zur Speicherung von Zahlen (Fließkomma, Integer etc.) und Datumsinformationen werden internationale Standards zur Darstellung (z.B. Norm IEEE 754 für Fließkommazahlen, Orientierung an ISO 8601 für die Datumsdarstellung) verwendet.

##### **4. Zugänglichkeit / Migrierbarkeit :**

Bei digitalen Repräsentationen sind einfache Kodierungen den technisch optimierten, aber komplexeren vorzuziehen (eher Rasterbilder als komprimierende Verfahren für Grafiken, eher Pulscode-Modulation mit linearer Quantisierung für Audioquellen etc.).

Digitale Formate, in denen die eigentliche Nutzinformation einfach und unmittelbar hinterlegt ist, können einfacher in neue Formate migriert werden und sind offener für eine "digitale Archäologie".

Kompression widerspricht der Anforderung der unmittelbaren Lesbarkeit. Wo aber zum Zeitpunkt der Erzeugung einer Quelle aus Praktikabilitätsgründen auch das Rohmaterial nur komprimiert bereitsteht (z.B. Video — MPEG2), wird nach Möglichkeit nur eine Version archiviert, die keine (oder möglichst wenige) Verluste beinhaltet (im Falle von Videodaten, also z.B. das Masterband).

Formate sind nicht von spezieller Hardware (z.B. PC-Plattform, spezielle Zusatzhardware wie Joystick oder Mikrophon, Dongles o.ä.) oder Software (z.B. nur spezielles Betriebssystem) abhängig.

### **5. Explizite Struktur / Selbstdokumentation :**

Formate, die sich selbst dokumentieren sind einfacher zu verstehen und daher weniger anfällig für vollständigen Informationsverlust durch Verlust von Dokumentationsunterlagen.

Auch dokumentenextern abgelegte Strukturbeschreibungen, die implizit aber Teil des Dokuments sind und nur zum Zwecke der Wiederverwendung, Mehrfachnutzung und Vereinheitlichung (z.B. Nutzung von DTDs in XML-Dokumenten) extern zum Zugriff bereit stehen, erfüllen dieses Bewertungskriterium. Hier ist aber das Kriterium 10 (Vermeidung von externen Dokumentenverknüpfungen) zu berücksichtigen.

Selbstdokumentierende Formate enthalten folgende Information:

Auszeichnung der eigenen technische Struktur (wie z.B. bei XML-basierten Verfahren),

Informationen zu Dokumentenversionen,

Informationen zur Integrität des Dokuments (z.B. Prüfsumme),

Inhaltlich beschreibende Metadaten (IPTC, EXIF, XMP o.ä.), im Archivumfeld ggf. angelehnt an Metadatenstandards wie Dublin Core oder OAIS-Prozessstandards,

Format- bzw. Medienspezifische Metadaten (Audio/Video: Lauflänge, Codec, Bandnummer etc.; Bild: Farbmodell, Fotodaten, Kameradaten etc.).

Bei Audiodaten: Einbindung und Umfang der Metadaten (Chunk)

### **6. Freie Nutzbarkeit (nicht proprietär)**

Die Nutzung eines Formates ist ohne Lizenzzahlungen möglich.

Formate, die Patente als Basistechnologien enthalten, können die Möglichkeit, dieses Format zu erhalten, hemmen oder sogar komplett verhindern. Nicht das Patent selbst als Schutz geistiger Urheberschaft ist das Problem, sondern dass nicht absehbar ist, wie die Patentinhaber in Zukunft damit umgehen werden.

Auch wenn für aktuelle Formate die Nutzung kostenlos oder sehr günstig möglich ist, verhindern Patente doch die Entwicklung von Open-Source-Software. Es ist außerdem nicht sichergestellt, dass nicht zu einem späteren Zeitpunkt die Nutzung mit möglicherweise hohen Lizenzgebühren verbunden sein kann.

Neuere ISO- oder DIN-Genormte Formate, die auf einer größeren Menge von Patenten beruhen (z.B. MPEG2/4 oder JPEG2000), sind weniger anfällig für dieses Patentproblem. Denn die Gefahr, dass alle Patentinhaber gemeinsam die Nutzung erschweren, ist deutlich weniger wahrscheinlich, als es bei einem einzelnen Patentinhaber wäre.

## **7. Vollständig und offen dokumentiert:**

Der Quellcode für Lesen und Schreiben des Formats ist als Open Source verfügbar.

Die freie Verfügbarkeit einer vollständigen Dokumentation ist sichergestellt.

Nicht-proprietäre, offene Standards sind üblicherweise besser dokumentiert und besser als proprietäre Formate.

Bei geschlossen definierten und dokumentierten Formaten sind die Formatdefinitionen (inkl. Quellcode und Tools) mindestens treuhänderisch in vertrauenswürdigen Archiven hinterlegt.

## **8. Verbreitungsgrad:**

Der Grad der Nutzung eines Formats durch die Erzeuger des Formats, durch beliebige Multiplikatoren oder auch durch alle potenziellen Nutzer von Informationsressourcen ist möglichst hoch.

Die reibungslose Austauschbarkeit über Systemgrenzen hinweg ist sichergestellt.

Vorteilhaft ist es, wenn ein Format bereits mit Systembordmitteln (Betriebssystem) oder mit regelmäßig verwendeten Werkzeugen (z.B. Browser) ohne weitere Plug-ins genutzt werden kann.

Die Unterstützung durch professionelle Werkzeuge der Formatsparten-Marktführer und eine breite Palette von vergleichbaren Werkzeugen, die das Format ebenfalls unterstützen, erhöht den Verbreitungsgrad deutlich.

Das Format wurde von anderen Archiven geprüft und als bevorzugtes oder unterstütztes Format beschrieben.

## **9. Stabilität / Reife:**

Die bevorzugte Version einer Formatdefinition ist eine finale Version, die sich bereits in der Praxis bewährt hat.

Das Format (oder die bevorzugte Version eines Formats) stellt nach Möglichkeit einen genormten Standard (IEEE, ISO o.ä.) oder einen De-Facto-Standard dar. Vom Format existieren nicht zu viele Versionen, die Versionen wechseln nicht zu häufig.

## **10. Nur bei Text- und Bildformaten: Keine Verknüpfbarkeit / Vermeidung von Interaktivität**

Die Vermeidung von notwendiger Interaktion zwischen elektronischen Unterlagen und externen Informationsressourcen oder von Benutzerinteraktion wird angestrebt.

Die Lesbarkeit und Verständlichkeit eines Dokument oder eines Datenbestandes erst auf der Basis von dokumenten-externen Informationsressourcen wird vermieden.

Je dynamischer oder interaktiver sich ein Datenbestand darstellt, umso höher wird der Aufwand für die Erhaltung, da in diesem Falle immer ein gewisser Anteil von Systememulation vorzuhalten ist, der die Interaktion ermöglicht.

## **11. Nur bei Audioformaten: Streamingfähigkeit**

Möglichkeit des Zugriffs auf das File über einen Streamingserver, ohne dass das gesamte File über einen möglicherweise sehr langwierigen Download geladen werden muss.



Beim Streamingserver ist jeder beliebige Punkt im Quellfile ansteuerbar.

## **12. Nur bei Audioformaten: Kodierungsverfahren**

Standardisiertes und offenes Kodierverfahren, zur Zeit nur PCM (Pulse Code Modulation) bekannt

## **4. Formatematrix zur Übersicht und Klassifizierung der Archivformate für Text-Bild- und Audiodokumente**

Download: [Formatematrix als Excel-Datei \(56 KB\)](#) (enthält 3 Tabellenblätter: Textformate, Bildformate, Audioformate).

Die verschiedenen zu untersuchenden Dateiformate wurden anhand der genannten Anforderungen und Kriterien bewertet. Dazu wurde ein Punkteraster von 0 für "sehr schlecht" bis 4 für "sehr gut" verwendet. Letztlich wird jedes Archiv selbst die Kriterien gewichten und die teilweise konkurrierenden Vor- und Nachteile der einzelnen Dateiformate im Hinblick auf die eigene Arbeit beurteilen müssen. Dafür kann in der Spalte "Gewichtung" (gelb hinterlegt) jedem der Kriterien die passende Prozentzahl zugeordnet werden, wobei die Summe der in dieser Spalte eingetragenen Prozentzahlen genau einhundert ergeben muss. Das Ergebnis ist dann in der dunkelgrün hinterlegten Spalte (3) ablesbar. Auf diese Weise fallen bei der Bewertung diejenigen Kriterien, die für ein Archiv von besonderer Bedeutung sind, auch besonders ins Gewicht. So kommt jedes Archiv zu spezifischen und nachvollziehbaren Entscheidungen über die Auswahl geeigneter Archivzielformate. Auf die Errechnung eines gewichteten Durchschnitts wurde hier verzichtet.

Die Bewertung der Dateiformate zeigt einen Arbeitsstand und muss, den Entwicklungen auf diesem Gebiet folgend, laufend ergänzt werden. Anhand einer Bewertungsmatrix, wie sie im Anhang aufgeführt ist, lassen sich die Dateiformate grundsätzlich in drei Gruppen klassifizieren:

### ***Bevorzugtes Archivzielformat/empfohlenes Archivierungsformat***

Sind bevorzugte Archivzielformate und erfüllen die hoch gewichteten Kriterien für Langfriststabilität in hohem Maß. Eine vollständige Erfüllung ist ausgeschlossen, weil sich z.T. Kriterien widersprechen.

Bei bevorzugten Archivzielformaten ist die Migrierbarkeit in möglicherweise später neu zu definierende Archivzielformate mit hoher Wahrscheinlichkeit sichergestellt.

### ***Akzeptierte Archivierungsformate***

Sind mögliche Dateiformate für die Archivierung elektronischer Unterlagen. Mittelfristig ist zumindest die Migrierbarkeit in später definierte Archivzielformate aus heutiger Sicht wahrscheinlich. Die Migrierbarkeit dieser Formate sollte aber regelmäßig geprüft werden.

Wenn ohne rechtliche Bedenken und mit vertretbarem Aufwand möglich, werden diese Formate schon bei der Archivierung in ein bevorzugtes Archivzielformat konvertiert.

Diese Dateiformate werden gewählt, wenn der Aufwand für Migration in ein bevorzugtes Archivzielformat zunächst als zu hoch angesehen wird oder wenn eine eindeutige Entscheidung zugunsten eines bevorzugten Archivzielformates noch nicht getroffen werden kann.

### ***Nicht akzeptierte bzw. umzuwandelnde Formate***

Sind nicht empfehlenswerte Archivierungsformate. Bei diesen Formaten kann schon kurzfristig die uneingeschränkte Nutzbarkeit der Nutzinhalt in Frage gestellt sein.

Für diese Formate wird empfohlen, schon bei der Archivierung eine Umwandlung in ein bevorzugtes Archivzielformat vorzunehmen.

## **Fazit: Archivare und IT sichern digitale Unterlagen langfristig**

Der Ansatz des AKEA zur Archivierung elektronischer Unterlagen ist praxisorientiert und für den Einsatz in allen Archivgattungen geeignet. Er ist als Handreichung konzipiert und Diskussionsgrundlage für das weitere Vorgehen. Erweiterungen sind unbedingt erforderlich, etwa im Hinblick auf die Archivierung von audiovisuellen und Zeichnungsdaten, von Unterlagen aus Dokumenten-Management-Systemen oder Intra- und Internetanwendungen.

Deutlich wird, dass die Archivare ihren Platz im Prozess zur Archivierung elektronischer Unterlagen finden müssen. Es mangelt ihnen nicht an Kompetenz, jedoch scheint zuweilen das Selbstbewusstsein, auch für diesen Bereich die Richtlinienkompetenz zur Archivierung wahrzunehmen, nur schwach ausgeprägt zu sein. Vielfach scheint es, als überließen die Spezialisten für die Archivierung diesen Bereich den Spezialisten für die Datenhaltung in der IT.

Es sind die Archivare, die der IT die Anforderungen für die langfristige Archivierung elektronischer Unterlagen benennen und die darauf dringen müssen, dass diese Anforderungen künftig selbstverständlich berücksichtigt werden, denn auch die Archivierung digitaler Unterlagen gehört in den Verantwortungsbereich der Archive.

---

### Anmerkungen

(1) Die im vorliegenden Artikel zusammengestellten Empfehlungen sind das Ergebnis von sechs Arbeitskreissitzungen zwischen November 2004 und Februar 2007 und wurden von den Mitgliedern des Arbeitskreises gemeinsam erarbeitet. Mitglieder sind: Jana Behrendt (SWR Baden-Baden), Dr. Martin Burkhardt (Wirtschaftsarchiv Baden-Württemberg), Thomas Dahm (Abbott GmbH & Co KG), Kai Franke und Antje Scheiding (Bertelsmann AG), Katja Glock und Bodo von Eberstein (Merck KGaA), Andreas Gaul (Dresdner Bank), Dieter Gross (Dr. Ing.h.c.F. Porsche AG), Dr. Ulrike Gutzmann (Volkswagen AG), Bettina Hasselbring (Bayerischer Rundfunk), Rainer Heid und Roman Sticht (BMW Group), Ulrich Kamp und Urs Lambertz (Aktion Mensch e.V.), Ulrich Kirchner und Heiko Keunike (Franz Haniel & Cie. GmbH), Dr. Christian Keitel (Landesarchiv Baden-Württemberg), Jürgen Klack, Dr. Annerose Knoch, Walburga Baier (Deutsche Post World Net), Britta Leise (Georg-Fischer AG+GF), Vladimir Marek (Degussa AG), Ulrich Melk (BPW Bergische Achsen Kommanditgesellschaft), Romy Meyer (Stadtarchiv Braunschweig), Elke Pfnür (Bayerische Hypo- und Vereinsbank AG), Dr. Christoph Popp (Stadtarchiv Mannheim), Dr. Stefan Przigoda (Deutsches Bergbau-Museum Bochum), Kornelia Rennert (Mannesmannröhren-Werke GmbH), Wolfgang Richter (Deutsche Telekom), Ute Schiedermeier (Siemens AG), Elke Sonnenberg (Robert Bosch GmbH), Dr. Hans-Jürgen Sproß (DaimlerChrysler AG), Sophie Wego (E.ON AG), Klaudia Wehofen (Stadtarchiv Düsseldorf), Manfred Witt (IHK Nordrhein-Westfalen). Besonderer Dank gebührt den Referenten der vergangenen Sitzungen: Dr. Matthias Kaack (Volkswagen AG), Prof. Dr. Siegfried Hackel und Tobias Schäfer (Physikalisch Technische Bundesanstalt Braunschweig), Wolf Zimmer (Ploenzke AG), Martin Angres (Media TV Angres GmbH), Mary Ellen Kitchens und Udo Appel (Bayerischer Rundfunk).

(2) Die Ergebnisse einschließlich der drei Matrizen zu Formaten für die Archivierung von Text- Bild- und Tondokumenten finden sich auch auf Homepage der Vereinigung deutscher Wirtschaftarchivare e.V unter <http://www.wirtschaftsarchive.de/akea/handreichung.htm>

(3) Spalte P bei Textformaten, Spalte O bei Bildformaten und Spalte Q bei Audioformaten