

nestor / Institut für Museumsforschung

Nicht von Dauer



Kleiner Ratgeber für die Bewahrung
digitaler Daten in Museen

2. Auflage

Materialien aus dem Institut für Museumsforschung – Sonderheft 2

Die Reihe „nestor – ratgeber“ wird herausgegeben von

nestor Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit
Digitaler Ressourcen in Deutschland

Network of Expertise in Long-Term Storage of Digital Resources

<http://www.langzeitarchivierung.de>

Die vorliegende Veröffentlichung erscheint in gemeinsamer Herausgeberschaft mit

IfM Institut für Museumskunde der Staatlichen Museen zu Berlin, Stiftung
Preußischer Kulturbesitz

<http://www.smb.spk-berlin.de/ifm>

© 2009 nestor/IfM

Der Inhalt dieser Veröffentlichung darf vervielfältigt und verbreitet werden, sofern der Name der Rechteinhaber "nestor - Kompetenznetzwerk Langzeitarchivierung" / IfM genannt wird. Eine kommerzielle Nutzung ist nur mit Zustimmung der Rechteinhaber zulässig.

Autoren dieser Veröffentlichung:

Dr. Stefan Rohde-Enslin, Institut für Museumsforschung (SMB-PK)

Dr. Keith Allen

Fotos:

Dr. Stefan Rohde-Enslin, Joshua Enslin

ISSN 1860-4641 (Materialien aus dem Institut für Museumsforschung. Sonderheft)

ISSN 1860-4706 (nestor-Ratgeber)

URN <http://nbn-resolving.de/urn:nbn:de:0008-2010030200>

Inhaltsverzeichnis

Vorwort zur zweiten Auflage	2
Liebe Kolleginnen, liebe Kollegen,	4
Wie man es sieht	6
Dateiformate	13
Formate im Einzelnen.....	15
1. Bildformate.....	15
TIF.....	15
JPG.....	17
JPEG 2000	19
GIF	20
WEITERE BILDFORMATE.....	21
2. Textformate	22
DOC.....	22
RTF	24
TXT	25
PDF.....	26
PDF/A	27
3. Videoformate	29
4. Audioformate	30
5. Datenbankdateien	31
6. Formate: Werkzeuge, etc.....	32
Speichermedien	33
Medien im Einzelnen.....	36
Diskette	36
ZIP-Diskette	38
Magnetband	39
Festplatte	41
CD.....	42
DVD.....	47
Blu-ray Disc.....	48
USB - Sticks	49
Zusammenfassung: Speichermedien	50
Weitere Überlegungen und Empfehlungen	53
Anhang	55
Übersicht der Empfehlungen	55
Keine Angst vor XML.....	56
E-Mail-Archivierung.....	58
nestor und mehr	63

Vorwort zur zweiten Auflage

Die erste Auflage dieser kleinen Schrift erschien 2004. Im Einklang mit dem Thema konnte auch ihr Inhalt nicht anders sein als „Nicht von Dauer“. Das darin behandelte Problemstellung, die Bewahrung digitaler Daten in Museen und anderen Kultureinrichtungen, ist nach wie vor vorhanden - eine einfache Lösung ist auch fünf Jahre später nicht in Sicht. Dennoch gab es eine Vielzahl von Entwicklungen in den letzten Jahren, die es zu beachten gilt. Die zunehmende Verbreitung von PDF/A und das Aufkommen von Flash-Speicher-Medien seien hier nur stellvertretend genannt.

Die Neuauflage dieser Broschüre greift einige dieser Entwicklungen auf und wägt sie in Hinblick auf die Problemstellung ab. Beim Überarbeiten wurde deutlich, dass die Umsetzung von Neuerungen sehr Ungleichmäßig von statten geht. So wurde der Gedanke, auf das Kapitel über Disketten in der Neuauflage zu verzichten, zurückgenommen, weil klar wurde, dass sie noch vielfach in Gebrauch sind. Die einen bleiben beim Altbekannten während die anderen stets auf der Suche nach Neuerungen sind und längst ihre DVD-gegen Blu-Ray-Laufwerke ausgetauscht haben. Das eine wie das andere Verhalten ist – in Blick auf die Langzeitarchivierung – nicht unproblematisch.

Jener Teil, der sich grundsätzlich mit der Problematik und dem Umgang mit Neuerungen befasst, ist auch in der Neuauflage zu finden. An den allgemeinen Empfehlungen, die darauf hinauslaufen, Veränderungen offen gegenüber zu stehen, aber nicht sogleich jedem Trend zu folgen, hat sich wenig geändert. Ebenso an den Schlussfolgerungen: Langzeitarchivierung kann nur der erreichen, der sie sich zur beständigen Aufgabe macht!

Bereichert wird diese Auflage durch einen Anhang zur e-mail-Archivierung der etwas aus dem Rahmen fällt. In dem kurzen Text geht es weniger um konkrete Empfehlungen für die Archivierung als vielmehr um die zusammenfassende Darstellung der Komplexität und darum, zu zeigen, dass rund um den Erdball an Lösungen gearbeitet wird. Die Fragestellungen sind letztlich überall ähnlich. Auch deshalb haben wir uns entschlossen, diese Auflage ebenfalls in englischer Sprache zu veröffentlichen. Die Überarbeitung und Erweiterung entstand sozusagen in deutsch-amerikanischer Koproduktion (Stefan Rohde-Enslin und Keith Allen).

Ein nicht unwesentlicher Grund für die Neuauflage besteht schließlich darin, dass die erste Auflage vollständig vergriffen, die Nachfrage an gedruckten Informationen aber nach wie vor sehr groß ist.

Es wäre schön, wenn keine zukünftige Auflage dieses „Kleinen Ratgebers“ mehr nötig wird, weil das Problem der Bewahrung digitaler Daten grundsätzlich und für alle gelöst ist. Wenn sich also das Problem als ebenfalls „nicht von Dauer“ herausstellt. Viele arbeiten daran, aber das wird noch etwas dauern ...

Berlin, 2009

Liebe Kolleginnen, liebe Kollegen,

eigentlich ist alles ganz einfach. Alles ändert sich. Altes wird durch Neues ersetzt, nur Weniges bleibt für lange Zeit erhalten. Eine der Aufgaben von Museen ist es, nicht alles sogleich dem Zerfall zu überlassen. Wir bemühen uns, Ideen, Techniken, Werkzeuge und vor allem die daraus resultierenden Objekte möglichst lange zu erhalten. Um das zu erreichen, wenden wir ebenfalls Ideen, Techniken und Werkzeuge an. Auch diese ändern sich und werden durch neue ersetzt.

Datenbanken ersetzen Karteikästen, Bilddateien ersetzen Papierabzüge von Objektfotos und Textdateien ersetzen Briefbögen. Das hat seine Vorteile, vieles geht schneller und problemloser vonstatten, Bild und Text lassen sich leicht zusammenbringen und viel größere Mengen von Objekten lassen sich viel effektiver verwalten. Es scheint, als werde durch die Digitalisierung unserer Werkzeuge eigentlich alles ganz einfach, oder? Es lohnt sich genauer hinzuschauen.

Was sich ändert, ist nicht wenig. Es geht um mehr als das Ersetzen eines Karteikartenformats durch ein anderes oder des einen bevorzugten Filmmaterials durch ein anderes. Die Veränderung ist viel grundlegender. Immer mehr von dem, was wir für unsere Arbeit benötigen, liegt nur noch digital vor. Die Vorteile wurden genannt, sie sind immens. Die Nachteile dürfen jedoch keinesfalls übersehen werden. Wir werden in unserer Arbeit von Computern abhängig: Informationen von Karteikarten ließen sich mit bloßem Auge lesen, die Informationen aus Datenbanken dagegen können wir ohne Technik nicht abrufen. Versuchen Sie es: Stellen sie sich einfach vor, der Computer, mit dem Sie arbeiten, sei nicht mehr da. Welche Literaturverweise hatten Sie sich zum Objekt mit der Inventarnummer 1014 vermerkt?

Jetzt noch einmal zum Ausgangspunkt: Alles ändert sich. Schon in wenigen Jahren wird der Computer, mit dem Sie arbeiten, nicht mehr da sein! Die Entwicklung der Technik ist rasant und der Computer, der ihnen dann zur Verfügung stehen wird, wird wahrscheinlich ein ganz anderer sein als der, den sie kannten. Denken Sie fünf Jahre zurück. In der Informationstechnik hat sich seitdem einiges getan. Systemarchitekturen, Chips, Speichermedien, Betriebssysteme, Programme, ... und vieles mehr haben sich geändert. Es scheint gewiss, dass in weiteren fünf Jahren dies alles noch einmal anders sein wird. Noch weiter gedacht: Wie ist es nach zehn, fünfzehn oder zwanzig Jahren? Anders und wieder anders. Und dann erneut die Frage: Welche Literaturverweise hatten Sie sich (vor vielen Jahren) zum Objekt mit der Inventarnummer 1014 vermerkt? Wird Ihnen der neue Computer zu einer Antwort verhelfen können? Die Wahrscheinlichkeit dafür ist gering, sehr gering!

Das Problem lässt sich leicht übersehen, tritt es doch erst in einigen Jahren auf. Doch es ist sehr ernst zu nehmen, denn mit der Möglichkeit, auf unsere Daten zuzugreifen, verlieren

wir auch all jene Informationen, die wir für unsere Arbeit brauchen und die wir für künftige Generationen aufbewahren sollten.

Denn höchstwahrscheinlich können die alten Dateien von den neuen Programmen nicht gelesen werden. Ebenso wahrscheinlich ist, dass die neuen Betriebssysteme keine Lesegeräte für die alten Speichermedien mehr betreiben können. Um unsere Daten herum hat sich eben alles geändert. Der einzige praktikable Ausweg besteht darin, die Daten ebenfalls zu ändern, sie stetig an neue Programme anzupassen und sie von einer Art Speichermedien auf die nächste zu übertragen.

Eine einmal geschriebene Karteikarte ließ sich in den Schrank stellen und noch nach zwanzig oder fünfzig Jahren unversehrt wieder finden. Anders die digitalen Daten und Dateien: Ihre Erhaltung verlangt beständige Aufmerksamkeit. CD-ROMs beispielsweise sollten alle zwei bis drei Jahre kopiert werden, sie werden mit der Zeit unlesbar. Mit jedem neu eingeführten Programm und jeder neu eingeführten Programmversion ist zu prüfen, ob die zuvor erstellten Dateien noch lesbar sind. Sie sollten gegebenenfalls in einem neuen Dateiformat gespeichert werden ... Tatenlos lassen sich elektronische Daten nicht erhalten.

Schon beim Erzeugen von Dateien werden die Weichen dafür gestellt, wie viel und welcher Aufwand hinterher für den Erhalt der Informationen aufzubieten ist. Wie Daten erhalten werden können und wie sich der Aufwand dafür reduzieren lässt, davon handeln die folgenden Seiten. Möglichst konkret soll gezeigt werden, was getan werden kann und muss.

Die Vorschläge der folgenden Seiten sollten alle beherzigen, die in irgendeiner Weise Dateien produzieren oder für Daten verantwortlich sind. Nur wenige von uns sind (Hobby-) Informatiker, deshalb: Lieber klare Angaben statt ausgefeilter Definitionen, Beispiele statt Theoreme und lieber ausführliche Begriffe statt Fachjargon. Alle sollen nach dem Lesen sagen können: Eigentlich ist (das) alles ganz einfach.

Wie man es sieht ...

Die Bewahrung digitaler Daten beginnt mit dem Speichern derselben. Sie ist weit mehr als das. Um das Problem zu verstehen, sollte sich jeder, der solche Daten erhalten will, zunächst fragen, was eigentlich passiert, wenn er speichert. Die Frage liegt nahe - doch ganz streng genommen ist sie falsch gestellt. Wir speichern ja nicht wirklich selbst, sondern weisen ein Programm an, Daten in Form von Dateien an einem bestimmten Ort zu speichern. Das ist keine Wortklauberei. Tatsächlich hängt es vom gewählten Programm ab, in welcher Form es die Daten zu einer Datei zusammenfasst. Die Programme erlauben uns einen gewissen Einfluss: Wir können eine Bild-, Text-, Video-, oder Audiodatei in diesem oder jenem Dateiformat speichern – aber nur, wenn das Programm es erlaubt. Andere Programme benutzen andere Dateiformate. Oft führt das zu „Missverständnissen“.

Speichern Sie beispielsweise ein Bild im PSD-Dateiformat mit dem Programm „Adobe Photoshop“ und versuchen Sie jetzt, diese Datei mit einem anderen Programm zu öffnen. In den meisten Fällen wird dies nicht gelingen. Das genannte Programm hätte aber auch erlaubt, die Datei im TIF-Dateiformat zu speichern. Um diese Datei zu öffnen, hätten Sie eines von sehr vielen verschiedenen Programmen wählen können. Die Programme hätten „verstanden“, welche Daten an welcher Stelle in der Datei wie zu interpretieren sind.

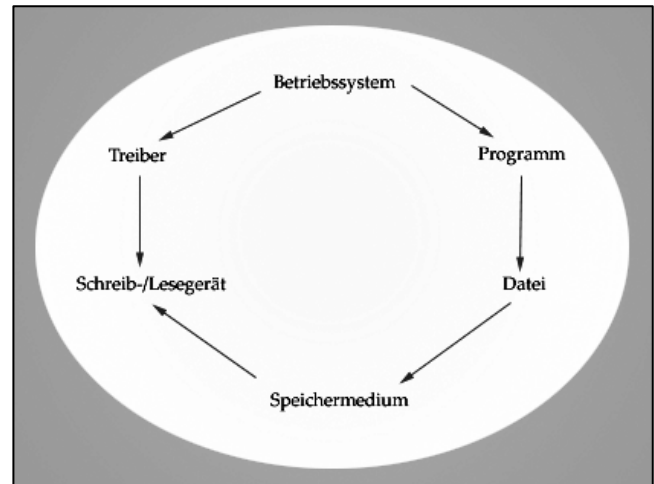
Zu Dateiformaten später mehr. Hier sollte vor allem festgehalten werden, dass Programme und Dateiformate eng verknüpft sind. Die meisten Programme können nur eine sehr begrenzte Anzahl von Dateiformaten lesen oder speichern. Angenommen, die Möglichkeiten der Dateiformatwahl, die Ihnen ein Programm bietet, reichen Ihnen nicht aus. Sie müssten das Programm wechseln. Doch welche Wahl hätten Sie? Nicht jedes Programm arbeitet unter jedem Betriebssystem! Bleibt die Möglichkeit, das Betriebssystem zu wechseln. Auch hier ist die Freiheit nicht grenzenlos, denn nicht jeder Rechner ist für jedes Betriebssystem geeignet. Immerhin können wir den Rechner frei wählen - die notwendigen Mittel vorausgesetzt.

Alles greift ineinander, bedingt sich gegenseitig. In dem Augenblick, in welchem wir speichern (oder das Speichern veranlassen), sind immer schon viele Entscheidungen gefallen – für den Rechner, das Betriebssystem, das Programm, das Speichermedium, das Aufzeichnungsgerät Im Alltag muss uns das nicht sonderlich interessieren - wenn wir Daten erhalten wollen, schon. Wichtig ist die Sichtweise:

Wer Daten speichert, benutzt dazu ein bestimmtes Programm, das unter einem ganz bestimmten Betriebssystem auf einem ganz bestimmten Rechnertyp arbeitsfähig ist. Mit diesem Programm speichert er seine Dateien in einem bestimmten Dateiformat auf einem bestimmten Speichermedium. Um Daten von einem Speichermedium des gewählten Typs

zu lesen, braucht er ein Lesegerät von einem bestimmten Typ, und dieses Lesegerät muss wiederum mit seinem bestimmten Betriebssystem zusammenarbeiten.

Lassen Sie uns, der Einfachheit halber, dem Kind einen Namen geben: Jede Datei hat ihre eigene *Entstehungsumgebung*, deren Komponenten wechselseitig voneinander abhängig sind. Lassen Sie uns auch - wo wir schon dabei sind - die Umgebung, in der die Datei aufgerufen wird, mit einem Namen versehen und von der Nutzungsumgebung sprechen. Die *Nutzungsumgebung* besteht aus Komponenten der gleichen Art, und auch hier sind die Komponenten wechselseitig voneinander abhängig.



Wenn wir eine Datei in einer Umgebung speichern und sie in derselben Umgebung gleich wieder aufrufen, dann sind Entstehungs- und Nutzungsumgebung identisch. Das geht problemlos. Wenn wir eine Datei speichern und anschließend versuchen, sie auf einem anderen Rechner zu öffnen, dann bringen wir die Datei in eine andere Umgebung. Die Wahrscheinlichkeit, dass wir die Datei lesen können, hängt davon ab, wie weit sich diese Umgebungen unterscheiden. Genauso ist es, wenn wir eine Datei speichern und sie einem Kollegen übermitteln, der sie öffnen und bearbeiten soll: Wir ändern die Umgebung und der Unterschied zwischen Nutzungsumgebung und Entstehungsumgebung bestimmt, wie groß die Wahrscheinlichkeit ist, die Datei nutzen zu können.

Dieser Blick auf zwei verschiedene Umgebungen zeigt, dass die Frage, ob wir unsere Dateien noch in einigen Jahren werden lesen können, von vielerlei Faktoren abhängt. Offensichtlich ist, dass Speichermedien verfallen: Disketten und CDs sind schon nach wenigen Jahren nicht mehr lesbar. Weniger offensichtlich ist, was häufiger Probleme bereitet: Die Dateien sind erhalten, lassen sich aber nicht mehr verwenden, weil es keine Programme mehr gibt, die sie interpretieren können oder die Dateien sind auf einem Medium gespeichert, für das es kein Lesegerät mehr gibt. Alte Programme und alte Lesegeräte verlangen alte Betriebssysteme ...

Jeder Unterschied in jeder einzelnen Komponente, der zwischen Entstehungsumgebung und Nutzungsumgebung einer Datei tritt, verringert die Wahrscheinlichkeit, diese Datei auch in Zukunft noch nutzen zu können.

Damit ist das Übel benannt: „Der Unterschied“ oder – aus zeitlicher Perspektive – „die Veränderung“. Sorgen wir dafür, dass es keine Veränderung gibt und wir haben das Problem gelöst! Ganz so einfach ist es leider nicht. Irgendwann wird auch der robusteste

Rechner nicht mehr funktionieren, irgendwann wird auch die langlebigste CD oder Diskette nicht mehr zu lesen sein und irgendwann wird es möglicherweise keine Programme mehr geben, die das von uns gewählte Dateiformat (und damit die Daten) lesen können. Die Veränderung lässt sich nicht verhindern! Müssen wir deshalb tatenlos zusehen, wie unsere Daten von ganz allein ihre Nutzbarkeit verlieren? Nein, denn wir können zwar die Veränderungen nicht wirklich verhindern, wir können jedoch Einfluss auf die Veränderung und vor allem auf die Auswirkungen der Veränderung nehmen.

Veränderungen mit Bedacht vornehmen und kontrollieren

Jede Veränderung in Soft- und Hardware ist eine Gefahr für die Existenz digitaler Daten. Wer plant, ein neues Programm zu kaufen, der sollte schon im Vorhinein prüfen, ob damit die alten Dateien noch lesbar sind. Genauso sollte jeder, der einen neuen Rechner kauft, prüfen, ob die Lesegeräte, die er für seine bisher benutzten Speichermedien braucht, mit diesem Rechner zusammenarbeiten. Was nützt dem, der bisher auf Disketten gesichert hat, ein neuer und schneller Rechner, wenn der nicht erlaubt, ein Diskettenlaufwerk zu betreiben und was nützt ein neues und schnelleres Betriebssystem, wenn sich damit die bisher benutzten Bandlesegeräte nicht betreiben lassen?

Vor jeder Veränderung der Umgebung, in der Dateien benutzt werden, ist zu prüfen, ob die neue Umgebung gestatten wird, aus den alten Dateien Informationen zurückzugewinnen, und nach jeder Veränderung dieser Umgebung ist zu prüfen, ob die alten Dateien tatsächlich noch lesbar sind.

Veränderungen, die sich vermeiden lassen, weil sie nur auf den ersten Blick Vorteile bringen, die alten Daten jedoch entwerten, sollten vermieden werden.

Veränderungen rechtzeitig vornehmen

Auf der anderen Seite gibt es viele gute Gründe für Veränderungen. Wer etwa plant, fortan verstärkt mit Videodateien zu arbeiten, der wird seinen betagten Rechner ersetzen müssen. Das ist nur eine jener Situationen, die unweigerlich eintreten, weil die Möglichkeiten der Technik ständig erweitert werden. Ein schwerwiegender Grund für Veränderung tritt spätestens dann ein, wenn der alte Rechner nicht mehr funktioniert. Ein allzu langes Zurückstellen von Veränderungen ist auch im Hinblick auf den langfristigen Erhalt digitaler Daten nicht ratsam:

Ein Programm speichert Dateien in einem bestimmten Format, das war schon festgestellt worden. Der Aufbau der Datei und die Menge und Art der Zusatzinformationen, das Format also, welches die nächste Version des Programmes benutzt, weicht aber oft von dem bevorzugten Format der Vorgängerversion ab. Mit der neuen Programmversion kommt häufig eine neue Formatversion! Das ist nicht immer erkennbar, weil die Pro-

grammhersteller meistens die Bezeichnung (etwa „DOC“) beibehalten. Kurzfristig ist das kein besonderes Problem, denn die neue Programmversion ist meistens so programmiert, dass sie Dateien, welche mit ihrer direkten Vorgängerversion erstellt wurden, lesen kann. Schon mittelfristig aber, wenn die übernächste Programmversion eingeführt wird, kann es zu ernsthaften Problemen kommen. Die neueste Programmversion kann die Dateien ihrer Vorgängerversion lesen – Dateien noch früherer Versionen sind ihr zumeist unbekannt. Grafisch lässt sich das folgendermaßen darstellen:

	Dateiformat 1	Dateiformat 2	Dateiformat 3	Dateiformat 4
Programmversion 1	XXX			
Programmversion 2	XXX	XXX		
Programmversion 3		XXX	XXX	
Programmversion 4			XXX	XXX

Tabelle: Lesbare Formate

Wer mit der Programmversion 1 Dateien im Dateiformat 1 geschrieben hat und dann auf die Programmversion 2 wechselt, der kann diese Dateien noch lesen und bearbeiten. Wechselt er aber gleich auf die Programmversion 3, bekommt er Schwierigkeiten, denn diese Version des Programmes kann noch Dateien des Dateiformates 2, nicht mehr aber solche des Dateiformates 1 lesen.

Der einzig praktikable Ausweg besteht darin, mit der Version 2 seine ursprünglichen Dateien (die im Dateiformat 1 vorliegen) zu laden und sie im Format der Programmversion 2 zu speichern. Dieses Dateiformat lässt sich von der Programmversion 3 lesen. Wird Programmversion 4 eingeführt, dann ist es ratsam, jene Dateien, die zuvor mit der Programmversion 2 ins Dateiformat 2 gebracht worden waren, erneut zu behandeln, d.h. sie in Programmversion 3 zu laden und in dessen Format zu speichern ...

Alle fraglichen Dateien müssen in diesem Beispiel (und oft auch in der Praxis) mit jedem Wechsel der Programmversion neu geladen und gespeichert werden. Sie müssen ständig auf der Hut sein, denn wenn Sie eine oder mehrere Neuerungen der Programmversion überspringen, können Ihre Daten möglicherweise nicht mehr gelesen werden! Sie dürfen, aus dieser Sicht, nicht zu lange mit Veränderungen warten, sonst sind die Zwischenversionen der Programme nicht mehr erreichbar.

Auch wenn Sie nicht die Programmversion, sondern das Programm selbst wechseln, sollten Sie alle ihre Dateien in das neue Programm laden und von diesem aus speichern. Die Wahrscheinlichkeit, dass auch das noch später folgende Programm die Dateien lesen kann (die eine oder die andere Formatversion – sie sollten beide bewahren) wird dadurch größer. Ein Programmwechsel erfordert das gleiche Maß an Aufwand und Mühe und die gleiche ständige Wachsamkeit wie ein Wechsel der Programmversion.

Ständiges Überführen von Dateien in neuere Formate birgt aber auch die Gefahr von Veränderungen, denn oft können neuere Programme (oder Programmversionen) ältere Dateien noch lesen, können sie aber nicht mehr vollständig interpretieren. Möglicherweise funktioniert die Platzierung von Fußnoten nicht mehr einwandfrei oder die Einbindung von Bildern und Grafiken bereitet nach der Umwandlung Schwierigkeiten. Mit jeder Umwandlung steigt die Wahrscheinlichkeit dafür, dass es gravierende Unterschiede zwischen der ursprünglichen Datei und der letzten Version gibt – das heißt aber nichts anderes, als dass Informationen verloren gegangen sind.

	Dateiformat 1	Dateiformat 2	Dateiformat 3	Dateiformat 4
Programmversion 1	XXX			
Programmversion 2	(XXX) →	YYY		
Programmversion 3		(YYY) →	ZZZ	
Programmversion 4			(ZZZ) →	AAA

Tabelle: Veränderung des Dateiinhaltes durch Migration

Aus diesem Tal der Mühen und Risiken gibt es keinen Ausweg. Die einzige Chance, die wir haben, besteht darin, gar nicht erst (oder möglichst wenig weit) hineinzugehen! Das ist recht einfach.

Die Auswirkungen von Veränderungen minimieren

Die Probleme der wechselnden Dateiformate rühren daher, dass die Kontrolle über die Gestaltung und Benennung des Formates, welches ein Programm benutzt, ausschließlich den Herstellern dieses Programmes obliegt. Nach eigenem Ermessen können sie Änderungen vornehmen. Und: Weil die neuere Programmversion neue Fähigkeiten mitbringt, die ja auch in den Dateien gespeichert werden müssen, werden die Formate ständig weiterentwickelt. Die Dateiformate, die von einem Hersteller kontrolliert werden, heißen proprietäre Dateiformate. Oft ist die Definition eines solchen Formates nicht offen gelegt, so dass die Hersteller anderer Programme dieses Format nicht berücksichtigen können.

Ein Beispiel: „Microsoft Word“ verwendet solch proprietäre, nicht offen gelegte Dateiformate. Das gleiche Programm bietet aber die Möglichkeit, Dateien im „Nur-Text“-Format zu speichern. Die entstehenden Dateien sind sehr klein, denn sie verzichten auf alle Zusatzinformationen, wie die verwendete Schriftart, Schriftgröße, Schriftfarbe ... Das ist ein Nachteil. Doch dieses Format lässt sich - und neben dem Gewinn an Speicherplatz ist dies der größte Vorteil - mit einer Vielzahl von Programmen lesen, mit neuen und alten Versionen. Das so genannte ASCII-Format ist seit den späten 1960er Jahren in Gebrauch. Die Wahrscheinlichkeit, dass es noch weit in die Zukunft von Programmen gelesen und interpretiert werden kann, ist groß. Die Vielzahl von Dokumenten, die bis jetzt in diesem

Format vorliegen und die Tatsache, dass jeder Hersteller es ohne Lizenzgebühren in sein Programm einbinden kann, sprechen dafür. Wann immer Sie auf besondere Textauszeichnungen (aber auch auf übersichtliche Gestaltung von Fußnoten, die Einbindung von Grafiken ...) verzichten können, sollten sie in diesem ASCII-Format (das oft die Endung „TXT“ trägt, s.u.) speichern.

Proprietäre Dateiformate zu verwenden bedeutet, sich abhängig zu machen. Was ist, wenn es diesen Hersteller in einigen Jahren nicht mehr gibt? Das Dateiformat ist entweder geheim oder als geistiges Eigentum geschützt. Besser ist es schon, Dateiformate zu verwenden, die zwar einem Software-Hersteller gehören, die aber dennoch offen gelegt sind und deren Nutzung anderen Herstellern freigestellt ist. Solche Dateiformate sind dann, wenn sie auch tatsächlich von einer Vielzahl von Programmen verwendet werden, relativ beständig: Viele Menschen werden mit vielen Programmen viele Dateien in diesem Dateiformat speichern und dies wird für die Hersteller zukünftiger Programme ein Anlass sein, die Möglichkeit, Dateien in diesem Format zu lesen und zu bearbeiten, in ihre Programme zu integrieren. Ein Beispiel hierfür ist das TIF-Format für Bilddateien – es gehört zwar der Firma Adobe, darf aber von allen benutzt werden. Was an welcher Stelle der Datei steht und wie es zu interpretieren ist, das ist für jedermann einsichtig, jeder darf Dateien dieser Art erstellen und sie mit der Bezeichnung „TIF“ versehen.

Eindeutig die beste Möglichkeit bieten Dateiformate, die weit verbreitet, offen gelegt und nicht das geistige Eigentum eines einzelnen Herstellers sind. Solche Formate sind das schon erwähnte TXT-Format für Texte, aber auch das JPG-Format für Bilddateien und das MPG-Format für Videodateien. Die beiden letztgenannten Formate wurden jeweils von einem Internationalen Konsortium geschaffen. Das MPEG-Konsortium (die Moving Picture Experts Group) etwa setzt sich aus 350 Herstellern und Universitäten zusammen. Seine Empfehlungen wurden von der International Standard Organization (ISO) übernommen. Ähnliches gilt für die Arbeit des JPEG-Konsortiums (der Joint Picture Experts Group). Auch ihre Vorschläge wurden zu internationalen Standards erhoben. Die Benutzung des Formates ist jedermann freigestellt und die Anzahl von Programmen mit denen sich entsprechende Dateien erstellen lassen, ist entsprechend groß.

Zu einzelnen Formaten später mehr, hier sollte vor allem festgestellt werden, dass sich die Auswirkungen von Veränderungen minimieren lassen, wenn statt eines besonderen und speziellen Dateiformates ein allgemein zugängliches und weit verbreitetes Dateiformat verwendet wird. Der gleiche Gedankengang und die gleiche Schlussfolgerung treffen auch auf Speichermedien zu. Je weiter verbreitet die Benutzung eines Speichermediums ist und je mehr Hersteller Geräte zum Lesen und Schreiben dieser Medien anbieten und vertreiben, umso größer ist die Wahrscheinlichkeit, dass es auch in einigen Jahren noch solche Medien sowie entsprechende Schreib- und Lesegeräte gibt. Spezielle Speichermedien, wie etwa ZIP-Laufwerke, sind vielleicht für die alltägliche Arbeit hervorragend geeignet, aus den genannten Gründen ist, wenn es um den Langzeiterhalt Ihrer digitalen Daten geht,

jedoch dringend abzuraten: Diese Laufwerke und Speichermedien werden nur von wenigen Herstellern produziert.

Aus der Perspektive der Langzeiterhaltung ist in allen der erwähnten Komponenten der Entstehungsumgebung digitaler Daten das Allgemeine dem Besonderen vorzuziehen. Extravagante Lösungen verschaffen vielleicht Vorteile für den Augenblick, auf lange Sicht sind sie hinderlich, denn irgendwann wird ein Transfer in ein gebräuchlicheres Dateiformat oder auf ein Speichermedium eines weiter verbreiteten Typs notwendig. Diese Arbeit und damit auch die Mühe der ständigen Wachsamkeit Veränderungen gegenüber lässt sich verringern, indem schon beim erstmaligen Speichern bewusst die entsprechenden Dateiformate und Speichermedien gewählt werden. Wie bereits gesagt: Die Bewahrung digitaler Daten beginnt mit dem Speichern derselben. Aber sie ist weit mehr als das.



Schleichender Datenverlust (auf herkömmliche Art)

Dateiformate

Über Formate wurde im vorhergehenden Abschnitt bereits festgestellt, dass alle Formate, die der ausschließlichen Kontrolle eines einzelnen Softwareherstellers unterliegen und deren Definition nicht offen gelegt ist, zu meiden sind. Weitverbreitete Formate sind dagegen zu bevorzugen. Bevor im Folgenden gesagt wird, welche einzelnen Formate aus Sicht der Langzeiterhaltung digitaler Daten zu empfehlen sind, soll noch eine grundsätzliche Überlegung angestellt werden.

Formate lassen sich einteilen in proprietäre und allgemeine Formate, das ist bereits geschildert worden. Sie lassen sich nach dem Grad ihrer Verbreitung, in weit und weniger weit verbreitete Formate, einteilen. Eine andere Einteilung ist jene nach dem Verwendungszweck. Textdateiformate sind grundsätzlich verschieden von Bilddateiformaten und diese wiederum unterscheiden sich entschieden von Videodateiformaten ... Es lässt sich deshalb nicht eine einzige Empfehlung für alle Fälle aussprechen.

Prinzipiell lassen sich Formate aber auch danach unterscheiden, ob sie die in ihnen enthaltenen Daten in komprimierter oder unkomprimierter Weise enthalten. Nahe liegt anzunehmen, dass für die Aufbewahrung über lange Zeiten hinweg, aufgrund der anfallenden Menge an Daten und Dateien, komprimierende Dateiformate zu bevorzugen sind – schließlich wird die Dateigröße durch Komprimierung zum Teil erheblich reduziert. Bisweilen, zum Beispiel bei Videodateien, ist Komprimierung, das Zusammenfassen der Daten nach einem ganz bestimmten Algorithmus, notwendig, weil die Datenmenge sonst die Möglichkeiten heute üblicher Rechner weit übersteigen würde. Dies steht außer Frage und lässt sich nicht umgehen. Wie aber ist es dort, wo die Wahl offen steht, ist es wirklich sinnvoll, Dateien, die für lange Zeit lesbar sein sollen, zu komprimieren?

Ein entscheidender Grund spricht gegen das Komprimieren:

Computer arbeiten im Allgemeinen mit nur zwei Zuständen, ja und nein oder eins und null. Digitale Daten werden entsprechend gespeichert: Durch kleine Erhöhungen (Ja) oder deren Abwesenheit (Nein) auf einer CD oder DVD oder durch kleine Magnetfelder in bestimmter Ausrichtung (Ja) oder deren Abwesenheit (Nein) auf einer Diskette oder einem Magnetband. Auch innerhalb des Rechners existieren nur diese zwei Zustände. Alles, was dem Rechner an Informationen über eine Tastatur, ein Mikrophon, eine Kamera, ... zugeführt wird, wird auf diese Weise übersetzt oder kommt bereits übersetzt in den Rechner. Die Information, die gespeichert werden soll, ist nach der Übersetzung nur eine Menge von Einsen und Nullen. Das menschliche Auge kann digitale Daten ohne Hilfsmittel weder sehen noch sind sie unmittelbar verständlich. Die Informationen, die bewahrt werden sollten, sind verschlüsselt. Das Wiedergewinnen der Informationen bedeutet dementsprechend die Entschlüsselung der digitalen Daten.

Damit diese Entschlüsselung, die Rekonstruktion, gelingen kann, muss zwingend der beim Speichern verwendete Schlüssel bekannt sein. Was nun geschieht bei der Kompression: Die ohnedies verschlüsselten Daten werden ein weiteres Mal verschlüsselt! Um die Informationen zu gewinnen, müssen demgemäß zwei (oder mehr) Schlüssel bekannt sein. Leicht einsehbar: Je mehr Schlüssel man hat, umso größer ist die Wahrscheinlichkeit, einen zu verlieren. Das spricht eindeutig gegen die Kompression. Es gibt keinerlei Sicherheit, dass auch in etlichen Jahren noch Programme existieren (und lauffähig sind), welche die Kompression rückgängig machen können!

Der letzte Absatz sprach gegen Kompression im Allgemeinen, wie sie etwa in ZIP- oder RAR-Formaten gespeichert wird. Neben dieser Art der Kompression, die – für den Augenblick – die Dateien zwar verschlüsselt, aber dennoch mit den vollständigen Informationen erhält, gibt es Kompressionen, welche darauf beruhen, Teile der Informationen aus den Dateien herauszufiltern. Ein solches verlustbehaftetes Komprimieren ist beispielsweise das Speichern von Bilddateien im JPG-Format. Klein, handlich, für den Bildschirm ausreichend und schnell durchs Internet zu übermitteln, hat dieses Format wichtige Vorteile. Langfristige Erhaltung lohnt sich hingegen nicht, denn die Erstellung solcher Dateien bedeutet letztlich nichts anderes als die Vernichtung eines Teiles der Informationen. Immerhin, das Format ist weit verbreitet und sein Kompressionsalgorithmus ist offen gelegt und darf von jedem Hersteller und Nutzer verwendet werden. Wer beispielsweise eine Internetseite, die mit solchen Dateien bestückt ist, für lange Zeit speichern will, der hat keine Wahl. Wer aber zum Beispiel eine Objektfotografie in digitaler Form bewahren möchte, warum sollte der auf Informationen verzichten? Besser ist es, für den Langzeiterhalt eine Datei zu erstellen, die sämtliche Informationen enthält. In der Situation, in welcher eine solche Datei genutzt wird (z.B. wenn das Objektfoto in eine Internetpräsenz eingebunden oder als e-mail-Anhang verschickt werden soll), kann daraus stets eine komprimierte Datei in der jeweils nötigen Qualität und Größe geschaffen werden.

Nach diesen Überlegungen lässt sich die im letzten Kapitel aufgestellte Maxime, die das Allgemeine über das Spezielle stellte, erweitern: Im Sinne einer langer Nutzbarkeit sollten Daten möglichst vollständig in einer möglichst weit verbreiteten und einfachen Form gespeichert werden. Diese Form der Zusammenstellung der Daten sollte offen gelegt und ohne Einschränkungen für jedermann nutzbar sein.

An diesen Kriterien werden im Folgenden einige gebräuchliche Formate für verschiedene Verwendungszwecke gemessen.

Formate im Einzelnen

1. Bildformate

TIF

Beschreibung:

„TIF“ oder richtiger „TIFF“ ist die Abkürzung für „Tagged Image File Format“. Das ursprünglich von der Firma Aldus in Zusammenarbeit mit Microsoft und anderen Firmen entwickelte Dateiformat liegt mittlerweile in seiner sechsten Version vor. Das Copyright lag ursprünglich im Besitz der Firma Aldus, es ging jedoch in den Besitz von Adobe über, als beide Firmen im September 1994 verschmolzen. Die ursprüngliche Definition des Formates stammt aus den 1980er Jahren. Version 6.0 unterscheidet sich von Version 5.0 nur darin, dass hierin festgehalten wird, dass Adobe Besitzer der Definition ist. Die letzte große Änderung in der Formatdefinition geschah beim Übergang von Version 4.0 auf 5.0 im Jahre 1988.

Das TIF-Format gestattet sehr hohe Farbtiefen. Der Umfang der größtmöglichen Datei ist auf 2 hoch 32 Bytes begrenzt – eine Grenze im Gigabyte-Bereich, die wohl eher theoretisch ist. Innerhalb dieses Formates sind mehrseitige Bilder erlaubt („multipage“). Auch ist es möglich, Dateien anderen Formates innerhalb des TIF-Formates zu speichern (beispielsweise „JPEG in TIF“). Viele Programme geben mit dem Speichern im TIF-Format zugleich die Möglichkeit, die Speicherung in einer Komprimierung (gebräuchlich: LZW-Kompression) vorzunehmen.

Kontrolle:

Eine einzelne Firma (Adobe) kontrolliert die Definition, jedoch nicht die Anwendung des Formates: Es gibt keinerlei Einschränkungen oder Zwang zur Entrichtung von Lizenzgebühren. Die Definition des Formates ist jedermann zugänglich. Jeder Software-Hersteller kann auf ihrer Grundlage entsprechende Programme erstellen. Das TIF-Format ist flexibel gehalten und Softwarehersteller dürfen eigene Varianten entwickeln.

Verbreitung:

Das Format ist sehr weit verbreitet, die Anzahl von Programmen, die TIF-Dateien lesen und schreiben können, ist unüberschaubar groß.

Vor-/Nachteile:

Eindeutiger Nachteil des Formates ist die Größe der entstehenden Dateien. Der Vorteil von TIF-Dateien besteht darin, dass alle Informationen der Bilddatei gespeichert werden.

Die Vielzahl von Programmen, die TIF-Dateien bearbeiten können, ist ein weiterer Vorteil, ebenso die Ausnutzung eines sehr großen Farbraumes.

Abwägung:

Das reine TIF-Format ist für die Langzeitarchivierung geeignet. Auf Kompression sollte jedoch verzichtet werden, ebenso auf das Speichern anderer (Unter-) Formate in TIF und auf die Aufnahme mehrerer Bilder in einer Datei.

JPG

Beschreibung:

Die gebräuchliche Bezeichnung „JPG“ bezeichnet nicht das Format, sondern die Organisation, die es geschaffen hat, die „Joint Photographic Experts Group“ (JPEG). Der von dieser Gruppe festgelegte Name des Formates lautet JFIF (JPEG-File Interchange Format). Die letzte Überarbeitung der Definition des Formates stammt aus dem Jahre 1992. Die Definition wurde schon zuvor, im August 1990, von der International Standards Organization (ISO) in den Rang einer Norm erhoben (ISO/IEC IS 10918) und der Gebrauch von JPG-Dateien wird von der International Telecommunication Union (ITU, als Empfehlung T.81) empfohlen. Dieser so genannte „baseline“-Standard erlaubte nur verlustbehaftet komprimierte Dateien. Später wurde er erweitert und als JPEG-LS Standard (ISO/IEC IS 14495-1 | ITU-T Recommendation T.87), welcher auch das Speichern unkomprimierter Dateien erlaubt, fortgeschrieben. Teile dieses neuen Standards unterliegen neben anderen einem Patent von Hewlett-Packard. Die Benutzung ist jedoch ausdrücklich jedermann freigestellt. Gleiches gilt für die Erweiterung des Standards, welcher die Dateien in einer Weise speichert, die es gestattet, die Bilder aus den Daten schrittweise zu rekonstruieren („Progressive JPG“).

Hohe Farbtiefen lassen sich im JPG-Format speichern. Der Grad der Kompression kann im Allgemeinen beim Erstellen der Dateien gewählt werden. Je größer die Kompression, desto größer der Informationsverlust – es entstehen die so genannten JPEG-Artefakte, quadratische Flächen ähnlicher Farben innerhalb des dargestellten Bildes.

Kontrolle:

Das JPG-Format wird von einer großen Gruppe von Firmen und Universitäten in Zusammenarbeit mit Internationalen Organisationen für Standardisierung und Kommunikation betreut. Die „baseline“-Definition ist patentfrei, ihre Nutzung unterliegt keinen Einschränkungen. Dateien, die auf Grundlage des späteren Standards (JPEG-LS) erstellt werden, d.h. Dateien verlustfreier Kompression, unterliegen dagegen mehreren Patenten. Dennoch wurde die Verwendung der patentierten Algorithmen ausdrücklich jedermann erlaubt.

Verbreitung:

Das JPG-Format ist vor allem im Internet und als Speicherformat von Digitalkameras weit verbreitet.

Vor-/Nachteile:

Der Vorteil des JPG-Formates liegt in der geringen Größe der Dateien. Der Nachteil besteht in der verlustbehafteten Speicherung.

Abwägung:

Im Sinne der Langzeiterhaltung ist dieses Format nur eingeschränkt zu empfehlen. Für eine Empfehlung sprechen die weite Verbreitung und die Offenlegung der Definition. Eindeutig gegen eine Empfehlung spricht aber, dass nur die Speicherung in verlustbehafteter Komprimierung tatsächlich patentfrei ist. Die Langzeiterhaltung von JPG-Dateien bedeutet mithin zugleich die Erhaltung dieses Verlustes. Dort, wo es möglich ist, sollte in jedem Falle das TIF-Format bevorzugt werden.

Sollte – aus welchen Gründen auch immer – das JPG-Format gewählt werden, so sollte auf eine Speicherung im progressiven Format oder in der so genannten verlustfreien Komprimierung verzichtet werden – beides unterliegt Patenten, die möglicherweise irgendwann zu Lizenzzwängen führen.

$$\begin{aligned} F(u, v) &= \frac{\Lambda(u)\Lambda(v)}{4} \sum_{i=0}^7 \sum_{j=0}^7 \cos \frac{(2i+1) \cdot u\pi}{16} \cdot \cos \frac{(2j+1) \cdot v\pi}{16} \cdot f(i, j) \\ \Lambda(\xi) &= \begin{cases} \frac{1}{\sqrt{2}} & \text{for } \xi = 0 \\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

Anwendung der diskreten Kosinus-Transformation bei der Kompression im JPG-Format

JPEG 2000

Beschreibung:

Die Nachteile des JPG - Formates (einige wurden hier bereits beschrieben) veranlassten die Joint Picture Experts Group ein neues Format, JPEG 2000, zu entwickeln. Die Formatdefinition wurde im Jahre 2001 als ISO Standard 15444 veröffentlicht. Durch die Verwendung eines anderen Kompressionsalgorithmus werden im neuen Format die so genannten JPG-Artefakte vermieden. Bilddateien vergleichbarer visueller Qualität sind wesentlich kleiner als im herkömmlichen JPG – Format. JPEG 2000 erlaubt zudem verlustfreies Speichern, d.h. Speichern ohne die Bildqualität zu beeinträchtigen.

Kontrolle:

Einige Teile der Format – Definition unterliegen dem Patentschutz. Die meisten Teile dieser Definition können dennoch ohne Einschränkungen verwendet werden. Um JPEG 2000 Bilddateien anzuschauen müssen die entsprechenden Programme mit zusätzlichen Funktionalitäten ausgerüstet werden.

Verbreitung:

Auch wenn einige größere Kultureinrichtungen, wie etwa die London's Wellcome Trust Library, JPEG 2000 verwenden, hat dieses Format noch keine weite Verbreitung gefunden. Die Softwarehersteller waren bisher sehr zögerlich dabei, ihre Programme mit den nötigen Funktionalitäten zu versehen.

Vor-/Nachteile:

Hinsichtlich der Komprimierung ist JPEG 2000 wesentlich besser geeignet als das herkömmliche JPG. Die Fähigkeit verlustfrei zu komprimieren macht dieses Format prinzipiell für die Langzeitarchivierung interessant. Allein die sehr geringe Verbreitung des Formats spricht deutlich gegen eine Empfehlung.

GIF

Beschreibung:

Anfang Januar 1995 beschloss die Firma Unisys, Lizenzgebühren für die Verwendung des LZW-Komprimierungsalgorithmus, für den sie das Patent hielt, zu verlangen. Das GIF-Format (Graphics Interchange Format), das aus frühen Tagen des Internets datiert und von CompuServe und Unisys entwickelt worden war, wurde damit kostenpflichtig. Analoges gilt für die Komprimierung von TIF-Dateien, die von vielen Programmen auf Grundlage dieses Algorithmusses angeboten wird. In der Formatvariante 89a erlaubt das GIF-Format animierte Grafiken. Ende 2003 ist das Patent von Unisys in den USA ausgelaufen. Patentrechte in Deutschland liefen erst am 18. Juni 2004 aus. Als Reaktion darauf, dass sie nunmehr dafür Gebühren zahlen sollten, dass sie in ihre Programme die Möglichkeit einbauten, Dateien im GIF-Format zu benutzen, unterstützen immer mehr Softwarehersteller in ihren relevanten Produkten das PNG-Format (s. unten). Um ihre Internet-Seiten auch für alte Browser lesbar zu halten, wird dennoch weiterhin das GIF-Format von vielen Seiten-Programmierern genutzt.

Das GIF-Format erlaubt lediglich 256 verschiedene Farben. Es erlaubt die Darstellung transparenter Flächen und die Speicherung mehrerer Bilder in einer Datei, wobei die Bilder nach einem festzuschreibenden Zeitintervall wechseln.

Kontrolle:

Die Definition des Formates stammt von CompuServe und Unisys. Der im GIF-Format enthaltene Komprimierungsalgorithmus ist (war) patentiert. Der Streit um Lizenzgebühren, den es um dieses Format gab, ist beispielhaft - ebenfalls die Reaktion der Programmhersteller, auf andere Formate mit ähnlichen Eigenschaften auszuweichen.

Verbreitung:

Noch immer ist das Format weit verbreitet, weil seine Verwendung auch älteren Internet-Browsern erlaubt, die entsprechenden Seiten zu lesen.

Vor-/Nachteile:

Neben der Patentfrage ist ein Hauptnachteil bei der Verwendung des GIF-Formates die Begrenzung auf nur 256 verschiedene Farben. Ein deutlicher Vorteil ist die sehr geringe Dateigröße.

Abwägung:

Vom Speichern im GIF-Format ist vor allem deshalb abzuraten, weil die Besitzer der grundlegenden Algorithmen schon in der Vergangenheit gezeigt haben, dass sie bestrebt sind, ihre Stellung auszunutzen. Die Zahl der Softwarehersteller, welche das Format nicht mehr in ihren Programmen berücksichtigen, nimmt zu.

WEITERE BILDFORMATE

Aus der Unzahl von Grafikformaten seien einige der bekanntesten erwähnt. Sie alle sind für die Langzeiterhaltung von Dateien nicht geeignet, da sie im Besitz nur eines Softwareherstellers sind und oft auch nur von wenigen Programmen aus erstellt, gelesen oder bearbeitet werden können.

BMP	(Microsoft: Bitmap)
PSD	(Adobe: Photoshop Dateiformat)
PSP	(JSAC: PaintShop Pro Dateiformat)
UFO	(Ulead: PhotoImpact Dateiformat)
CPT	(Corel: PhotoPaint Dateiformat)
DNG	(Adobe: "Digital Negative")
PNG	(Portable Network Graphic)

Meiden Sie diese Formate für die Langzeiterhaltung Ihrer Bild- und Grafikdateien, es könnte sie morgen schon nicht mehr geben.

2. Textformate

DOC

Beschreibung:

Das DOC-Format ist das eigene Format von Microsoft Word. Es ist sehr weit verbreitet. In ihm werden neben Texten auch Textauszeichnungen (Fett, Kursiv, ...) gespeichert. Ebenso können DOC-Dateien Makros oder Grafiken enthalten. Das Format ist hochkomplex und sehr eng an das Textprogramm Word und das Betriebssystem Windows angelehnt. Mit der Erweiterung der Möglichkeiten seines Textprogrammes hat Microsoft auch das DOC-Format verändert – ohne die Kompatibilitäten sicherzustellen. Genau genommen ist deshalb nicht von dem oder einem DOC-Format zu sprechen, sondern von mehreren, die zum Teil untereinander nicht austauschbar sind. Dateien, die in einer alten Version des Formates (d.h. mit einer alten Version des Textverarbeitungsprogramms) gespeichert wurden, lassen sich nicht problemlos mit neuen Versionen des Programmes öffnen.

Mittlerweile gibt es wenigstens 7 verschiedene Versionen (1.0, 2.0, 6.0, 7.0/95, 8.0/97, 9.0/2000, 10.0/2002) - ungefähr alle zwei Jahre wird eine neue eingeführt. Der Tradition lieber eigene Regeln einzuführen als sich an allgemein Übliches zu halten ist Microsoft auch beim Nachfolger des DOC-Formates, XML-Dateien mit der Bezeichnung DOCX, treu geblieben. Mit dem Umstieg Microsofts vom DOC-Format auf DOCX wird es höchste Zeit, seine alten DOC-Dateien in ein anderes Format zu überführen.

Kontrolle:

Die alleinige Kontrolle über die Formatdefinition für DOC liegt bei Microsoft. Die Definition ist nur teilweise offen gelegt. Zwar gibt Microsoft die Definition auf Anfrage bekannt, verpflichtet die Anfragenden jedoch zu einer Geheimhaltung.

Verbreitung:

Das Format ist sehr weit verbreitet.

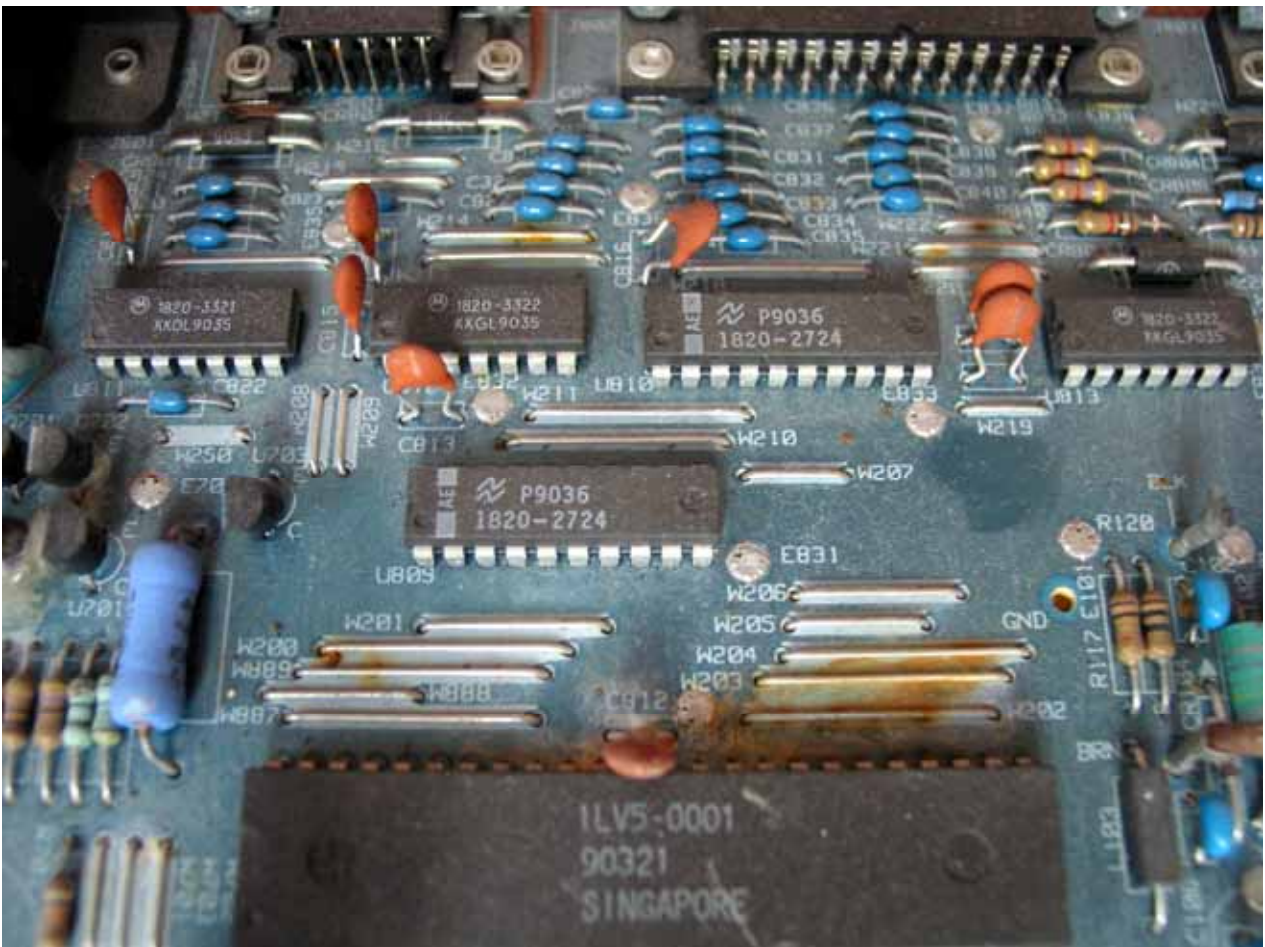
Weitere Vor-/Nachteile:

Recht umfangreiche Dateien entstehen auch dann, wenn tatsächlich nur Text (ohne besondere Auszeichnungen und in nur einer einzigen Schriftart und -größe) gespeichert wird.

Abwägung:

Die Lesbarkeit von Dateien im DOC-Format ist schon in der Gegenwart nicht sichergestellt, wenn es um die Weitergabe solcher Dateien geht: Selbst wenn der Empfänger einer DOC-Datei das entsprechende Microsoft Programm besitzt, kann er, bei einem Unterschied in der Programmversion, möglicherweise die fragliche Datei nicht benutzen. Hat der Empfänger ein anderes Programm als jenes, mit dem die fragliche Datei erstellt

wurde, bleibt ihm der Inhalt der Datei weitgehend unzugänglich. Im Sinne der Langzeiterhaltung ist das Format keinesfalls zu empfehlen, denn eine Weitergabe in die Zukunft potenziert die Wahrscheinlichkeit der Unlesbarkeit, die schon in der Gegenwart besteht.



Lebensdauer abgelaufen: Fundstück vom Straßenrand

RTF

Beschreibung:

Als Reaktion auf die Probleme der Inkompatibilität seines DOC-Formates über Programmversionen, Programme und Betriebssysteme hinweg hat Microsoft das RTF-Format eingeführt und sein Textverarbeitungsprogramm Word mit der Möglichkeit, in diesem Format zu speichern, ausgestattet. Dateien im Rich Text Format (RTF) lassen sich auch in anderen Programmen lesen. Allerdings ändert Microsoft die Definition von RTF mit beinahe jeder Version seines Textprogrammes. Die Austauschbarkeit ist dennoch vor allem deshalb größer, weil einerseits auf Versionskompatibilität geachtet wird (neue Programme können RTF-Dateien lesen) und andererseits das RTF-Format betriebssystemunabhängig ist. Dateien im RTF-Format basieren auf ASCII-Text (s.u. „TXT“) mit eingestreuten Steuerbefehlen.

Kontrolle:

Auch beim RTF-Format liegt die Kontrolle über die Definition ausschließlich bei Microsoft. Immerhin ist die Spezifikation offen gelegt und andere Softwarehersteller sind aufgerufen, ihre Programme mit der Möglichkeit auszustatten, RTF-Dateien zu lesen. Dennoch kann Microsoft jederzeit die Unterstützung von RTF (d.h. die Weiterentwicklung des Formates) einstellen oder in Richtung auf Inkompatibilitäten mit anderen als firmeneigenen Programmen verändern.

Abwägung:

Obwohl es wesentlich besser für eine Langzeitarchivierung von Text-Dateien geeignet ist als das DOC-Format, besteht auch beim RTF-Format die effektive Kontrolle durch nur einen Softwarehersteller. Aus diesem Grunde kann die Verwendung dieses Formates nicht empfohlen werden.

TXT

Beschreibung:

Das Kürzel „TXT“ steht an dieser Stelle für das Speichern im ASCII (oder besser ANSI) Format. In diesem Format werden die Textzeichen in einer einfachen Reihe hintereinander gespeichert. Textauszeichnungen sind nicht enthalten – fett markierter Text wird gespeichert, allerdings ohne die Markierungen. Gleichfalls können in diesem Format keine Querverweise oder eingebundene Grafiken gespeichert werden. Dieses Format ist mit nahezu jedem Textverarbeitungsprogramm les- und schreibbar. Die Dateiendungen variieren bisweilen. Manchmal lassen sich die Dateien als „TXT“, manchmal als „ASC“ speichern, manche Programme bieten das Speichern im „Plain Text“ an, was das Gleiche bewirkt. Das ASCII-Format ist die Basis sowohl für HTML, XML als auch für das RTF-Format.

ASCII steht für American Standard Code for Information Interchange. ASCII-Dateien werden seit 1963 verwendet. Schon bald nach Einführung des ASCII war eine Erweiterung nötig, denn es war ursprünglich nicht vorgesehen, internationale Zeichen zu berücksichtigen. Aus diesem Grunde wurde 1968 der erweiterte ANSI-Code (ANSI, American National Standards Institute) entwickelt, in welchem es nunmehr möglich war, beispielsweise dem deutschen Buchstaben „ö“ eine eindeutige Zahlenentsprechung zuzuordnen. Bei Microsoft (Word) heißt das Format „TXT“ und ist erreichbar unter „Speichern unter“ mit Angabe des Dateityps als „Nur Text“.

Kontrolle:

Genau genommen stellt der ASCII-Code nur eine jedem zugängliche Übersetzungstabelle dar. Das entsprechende Format besteht in der einfachen Aneinanderreihung. Weder diese Art der Reihung noch die Nutzung der Tabelle sind patentiert. Es gibt keinerlei Restriktionen der Verwendung.

Verbreitung:

Indirekt, d.h. als Grundlage für andere Formate, ist das ASCII-Format weit verbreitet. In direkter Form wird es selten verwendet. Aber nahezu jedes Textprogramm gestattet das Speichern und Lesen im ASCII-Format.

Weitere Vor-/Nachteile:

Entscheidender Nachteil des Formates ist die Unmöglichkeit, Textauszeichnungen (z.B. kursiv) und dem Textinhalt beigeordnete Objekte (z.B. Grafiken) zu speichern. Gewichtiger Vorteil ist die Kompatibilität über alle Betriebssysteme und fast alle Programme hinweg.

Abwägung:

Seit den frühen Tagen der Computer wird der ASCII-/ANSI-Code eingesetzt. Wann immer Sie auf Textgestaltungen verzichten können, sollten Sie dieses Format wählen.

PDF

Beschreibung:

Die Abkürzung PDF steht für Portable Document Format. Die Definition des Formates ist Eigentum des Softwareherstellers Adobe. Das Format entstand als Fortentwicklung einer Druckerbeschreibungssprache (Postscript) und wurde 1993 gemeinsam mit dem Adobe Acrobat, einem Programm zur Darstellung von Dateien dieses Formates, vorgestellt. Die Definition ist offen gelegt. Zum Lesen von PDF-Dateien stellt Adobe ein kostenloses Programm zur Verfügung, zum Erstellen der Dateien bietet der Hersteller ein kostenpflichtiges Programm an. Internet-Browser können durch kostenfrei erhältliche Zusatzprogramme PDF-Dateien direkt darstellen.

Kontrolle:

Die alleinige Kontrolle über das Dateiformat behält sich Adobe vor. Anders als beispielsweise beim DOC-Format von Microsoft ist die Definition jedoch offen gelegt und jedermann zugänglich. Den Herstellern von Programmen, die in der Lage sind, Dateien im PDF-Format zu speichern, werden keine Gebühren auferlegt. Dennoch bleibt das Format proprietär.

Verbreitung:

Das PDF-Format ist, vor allem der – in Relation zur Komplexität des Inhaltes – geringen Dateigröße und der Tatsache wegen, dass es kostenlose Leseprogramme gibt, sehr weit verbreitet.

Weitere Vor-/Nachteile:

Das PDF-Format entstand aus einer Druckerbeschreibungssprache. Eine klare Strukturierung der einzelnen Textteile ist in den Dateien nicht mehr erkennbar. Dies erschwert die Suche innerhalb von Dateien und die Transformation ihres Inhaltes in andere Formate.

Abwägung:

Wegen seiner gegenwärtig weiten Verbreitung ist das PDF-Format - wenn auch bedingt - für die Langzeiterhaltung geeignet, denn es ist einerseits ein Format, das von nur einem Hersteller kontrolliert wird (der aber anderen dieses Format zur Verfügung stellt) aber andererseits ist es ein recht kompliziertes, druckerorientiertes Format, dessen Rekonstruktion gegebenenfalls viel Aufwand erfordern wird.

PDF/A

Beschreibung:

PDF/A wurde als Antwort auf die Schwierigkeiten Textdateien über lange Zeit zu bewahren, geschaffen. Es handelt sich um eine spezielle Variante des PDF - Formats, das seit den 1990er Jahren breite Verwendung findet (s.o.). PDF/A-1a ist seit 2005 ein anerkannter ISO-Standard. Wichtig ist es, das allgemeine PDF - Format nicht mit PDF/A-1a gleichzusetzen. PDF/A-1a ist im Wesentlichen identisch mit der alten Version 1.4 des PDF-Formats und erlaubt z.B. nicht, Hypertextreferenzen oder Links, zu speichern.

Kontrolle:

Die Bemühungen einen Standard für die Langzeitarchivierung von elektronischen Texten zu finden gehen auf den Oktober 2002 zurück. So, wie mit seinem PDF – Format im allgemeinen, hat Adobe auch die dem PDF/A zugrunde liegende Formatdefinition öffentlich zugänglich gemacht und anderen die Weiterentwicklung gestattet. Es war eine Reihe großer Softwarefirmen die gemeinsam mit Kulturinstitutionen, wie z.B. der Library of Congress in den USA die Weiterentwicklung zu einem ISO - Standard betrieb. Die Library of Congress selbst hat dennoch zunächst mit der Umsetzung, d.h. der Empfehlung an die eigenen Mitarbeiter, dieses Format zu verwenden, gezögert, denn jedes Dokument das vorgibt, im PDF/A – Format vorzuliegen, ist zu prüfen, ob dies tatsächlich der Fall ist.

Verbreitung:

Die Verbreitung von PDF/A nimmt insbesondere in öffentlichen und kulturellen Einrichtungen schnell zu, vor allem seit es Validierungswerkzeuge (s.u.) gibt.

Weitere Vor-/Nachteile:

PDF/A-1, in seiner Form als ISO-Standard (19005-1) speichert Informationen zur Textdarstellung und weiteres in der gleichen Datei, d.h. die im Text verwendeten Schriftarten müssen unbedingt eingebettet werden. PDF/A-1 kann dennoch nur einen Teil der ursprünglichen Informationen (allerdings wesentlich mehr als beispielsweise das ASC-Format) aufnehmen und bewahren. Bemerkenswert ist, dass schon kurze Zeit nach der Einführung als ISO-Standard die Weiterentwicklung zu einer neuen Version begann. Ein Standard, der schon bald wahrscheinlich schon keiner mehr ist.

Abwägung:

Auch wenn eine Vereinfachung der Komplexität des PDF/A-Standards begrüßenswert scheint, sollte die Tatsache, dass fast im gleichen Augenblick in welchem der Standard verabschiedet wurde die beteiligten Institutionen begannen, einen Nachfolge - Standard zu entwickeln, ein Grund zum Nachzudenken sein. Dokumente, welche sie in PDF/A-1 speichern werden höchstwahrscheinlich auch in Zukunft lesbar (reproduzierbar) sein, das „Look and Feel“ wird wahrscheinlich erhalten bleiben. Dennoch: PDF/A-1 ist nicht die

Lösung für alle Probleme und vor allem ist es wichtig sicherzustellen, dass jene Dateien, die man in PDF/A-1 speichert auch tatsächlich der Definition gemäß gestaltet sind.

Mehr Information:

Eine Untersuchung der Königlichen Bibliothek der Niederlande (von 2007) zeigt einige der Probleme mit PDF/A auf:

http://www.kb.nl/hrd/dd/dd_links_en_publicaties/PDF_Guidelines.pdf

3. Videoformate

Um Videos ohne wahrnehmbares Ruckeln darstellen zu können, werden pro Sekunde wenigstens 25 Bilder angezeigt. Die entstehenden Datenmengen sind riesig. Auch wenn aus Sicht der Langzeiterhaltung digitaler Daten (und dessen, was sie repräsentieren) Komprimierungen prinzipiell abzulehnen sind, weil sie ein weiteres Verschlüsseln bedeuten (s.o.), kann im Videobereich nicht darauf verzichtet werden. Um so wichtiger ist es, einen Standard zu wählen, der weit verbreitet ist und von vielen Herstellern unterstützt wird, der zudem offen gelegt ist, so dass Dateien in diesem Format in späteren Jahren gegebenenfalls rekonstruiert werden können.

Die Motion Picture Experts Group (MPEG) hat seit ihrer Einrichtung 1988 mehrere solcher Standards entwickelt. Sie ist eine Arbeitsgruppe der International Standards Organization (ISO), in der viele führende Hersteller von Soft- und Hardware vertreten sind.

Während das AVI-Format von Microsoft und das MOV-Format von Apple stammen, ist das MPG-Format der Motion Picture Experts Group unabhängig von Betriebssystemen und Herstellern. Die gängigen Standards sind gegenwärtig:

MPEG1	Für Video-CD und für mp3-Dateien
MPEG2	Für DVD und digitales Fernsehen
MPEG4	Für Multimediaanwendungen
MPEG7	Für die Analyse und die Suche in Videos

Die Vorgaben der MPEG-Gruppe beschränken sich allerdings auf die Verwendung bestimmter Algorithmen. Softwarehersteller können darüber hinaus weitere Kompressionsalgorithmen in ihre eigenen Dateiformate einbauen. Gerade im Videobereich ist trotz der Einrichtung eines Standards die Vielzahl von tatsächlichen Dateiformaten unüberschaubar.

Die Empfehlung kann also nur dahin gehen, sicherzustellen, dass wenigstens die Festschreibungen der MPEG-Gruppe (als kleinster gemeinsamer Nenner) im gewählten Dateiformat berücksichtigt sind. Auch die darüber hinausgehenden Teile der Definition des Dateiformates sollten möglichst nicht von nur einer Firma kontrolliert werden.

AVI	Audio Video Interleaved – Format (Microsoft)
MOV	Apple Quicktime Format und
WMV	Windows Media Format (Microsoft)
FLV/SWF	Flash Video (Adobe Systems)

sind einige Beispiele für proprietäre Video-/Filmformate.

4. Audioformate

Ebenso wenig vereinheitlicht wie die Videoformate sind die gebräuchlichen Audioformate. Das schon 1987, auf Grundlage der von der MPEG-Gruppe geschaffenen Standards, eingeführte mp3-Format findet zwar immer weitere Verbreitung, hat sich jedoch noch nicht überall durchgesetzt. Es arbeitet mit Datenkomprimierung.

Das üblicherweise ohne Komprimierung arbeitende WAV-Format (von Microsoft und IBM) eignet sich nur für kleine Tonsammlungen oder solche mit ausreichend Finanzmitteln für Speicherplatz. Mit dem MID-Format lassen sich keine hohen Qualitäten realisieren, dafür sind die entstehenden Dateien sehr klein. Das von Apple geschaffene AIFF-Format ist wenig verbreitet und bietet höchstens Radioqualität.

Für Audioformate kann, in Hinblick auf die Langzeiterhaltung, derzeit keine eindeutige Empfehlung gegeben werden.

Was sich in Kultureinrichtungen (Museen, Archiven, Tonarchiven, ...) als Standard abzuzeichnen beginnt ist PCM WAV 96 khz / 24 bit. Dies ist auch eine Empfehlung der International Association of Sound and Audiovisual Archives (IASA).

5. Datenbankdateien

Datenbanken setzen sich aus mehreren Teilen zusammen. Kleine Programme, welche Formulare, Abfragen, Berichte oder ähnliches erzeugen, werden mit Hilfe eines Datenbankmanagementsystems geschaffen. Die Daten selbst sind üblicherweise in Tabellen zusammengefasst. Die einzelnen Dateien (für Verwaltung, Programme, Daten) werden entweder in einer Datei oder in mehreren Dateien gespeichert, wobei mehrere verschiedene Dateiformate verwendet werden können.

Anders als Bild- oder Textdateien, die zumeist etwas Vollständiges enthalten, sind Datenbanken in den meisten Fällen dynamisch konzipiert. Sie sind nicht abgeschlossen, sondern dazu angelegt, immer weitere Informationen aufzunehmen.

Für die Langzeiterhaltung der digital kodierten Informationen kann es deshalb immer nur um die Sicherung eines augenblicklichen Zustandes gehen. Eines Zustandes, der wenn möglich sowohl die verwendeten Programme (d.h. die Formulare, Abfragen, ...) als auch die Daten selbst (d.h. zumeist, die Tabellen) umfasst. Im Allgemeinen lässt sich das Speicherformat für die kleinen Programme nicht beeinflussen. Das Speicherformat für die Tabellen hingegen lässt sich bei vielen Datenbankmanagementsystemen wählen.

Wenn Ihr Datenbanksystem erlaubt, Daten im so genannten CSV-Format (CSV, Comma Separated Value) zu exportieren, dann sollten Sie von Zeit zu Zeit einen solchen Export vornehmen. Die einzelnen Einträge der grundlegenden Tabellen werden dabei im TXT (oder ASC-Textformat, s.o.) gespeichert, wobei sie durch Kommata getrennt werden.

6. Formate: Werkzeuge, etc.

Was ist überhaupt ein Format? Technisch gesehen besteht ein Format aus Regeln die es ermöglichen eine sinnvolle Beziehung zwischen dem Inhalt einer Datei (ein Text, ein Bild, ein Video, eine Reihe Töne, etc.) und dem Strom aus Einsen und Nullen, den der Computer allein versteht, festzulegen. Wenn das Ziel lautet, Bedeutung und Inhalt von Dateien, also mehr als nur lange Aneinanderreihungen von Einsen und Nullen zu erhalten, dann ist die Frage der Formatwahl von entscheidender Bedeutung. Werkzeuge und Informationsquellen (in englischer Sprache):

JHOVE

Werkzeuge werden gegenwärtig entwickelt, die helfen, mit Formaten umzugehen. Das erste heißt JHOVE (gesprochen: „jove“). Das JHOVE2 Projekt wird von der Library of Congress in den USA unterstützt. JHOVE soll zwei Fragen beantworten. Die erste lautet: „Ich habe irgendein digitales Objekt, welches ist sein Format?“. Wenn dieses geklärt werden konnte, kommt die zweite Frage: „Ich habe ein digitales Objekt, welches scheinbar in Format X vorliegt; ist es wirklich das Format X?“ JHOVE ist ein möglicherweise wichtiges Werkzeug, vor allem, wenn es um komplexe Formate geht, wie unter anderem PDF/A. Die Zukunft gehört möglicherweise digitalen Objekten in denen verschieden Dateien und Formate willkürlich kombiniert sind. Allein schon deshalb lohnt sich ein Besuch der Internetseite des Projektes. Sie finden dort weitere Empfehlungen.

<http://confluence.ucop.edu/display/JHOVE2Info/Home>

AONS

Das Automated Obsolence Notification System (AONS) informiert Sie, wenn Dateiformate unüblich geworden sind oder in der Gefahr stehen unüblich zu werden. Das Projekt wird von der Australischen Nationalbibliothek betrieben. Eine Beta-Version der Software finden sie hier:

<http://sourceforge.net/projects/aons/>

PRONOM

Das PRONOM-Projekt, betrieben vom Nationalarchiv Großbritanniens bietet sehr hilfreiche Informationen, insbesondere über Dateiformate im Zusammenhang mit Langzeitarchivierung.

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

Speichermedien

Unabdingbar ist das Problem der Langzeiterhaltung digitaler Daten mit der Verwendung geeigneter Speichermedien verbunden. Bewusst ist hier die Frage nach den Speichermedien hinter eine Betrachtung der geeigneten Dateiformate gestellt worden, denn – auch wenn es nahe liegt anzunehmen, die „Lebensdauer“ digitaler Daten sei in erster Linie genau so groß wie die Lebensdauer der Datenträger, so zeigt die Praxis, dass oftmals die Datenträger noch intakt sind, während die Dateien schon längst nicht mehr gelesen werden können, weil geeignete Programme fehlen. Das heißt aber keinesfalls, dass die Frage danach, welcher Datenträger für die Langzeitarchivierung geeignet ist, von untergeordnetem Interesse ist.

Die zur Verfügung stehenden Datenträger lassen sich grob in zwei Gruppen teilen: magnetische Datenträger, wie Magnetbänder oder Festplatten, und optische Datenträger, wie CDs und DVDs. Es gibt noch weitere Arten von Datenträgern, die sollen hier jedoch, ihrer geringen Verbreitung wegen, nicht weiter besprochen werden. Schon eingangs dieser Broschüre wurde festgestellt, dass unbedingt das Allgemeine dem Speziellen vorzuziehen ist. Ausgefeilte Speichermedien und die dazugehörigen Schreib- und Lesegeräte, können, wenn sie nicht weit verbreitet sind, jederzeit vom Markt verschwinden und damit die Anstrengungen zur Datenerhaltung zunichte machen. Zudem sind die allgemeineren Lösungen in zweierlei Hinsicht günstiger: Zum einen finanziell, denn ob der hohen Verbreitung ist zumeist der Preis für ein einzelnes Medium recht begrenzt, und zum anderen hinsichtlich des Datenaustausches für die Gegenwart, denn die Wahrscheinlichkeit, dass die Empfänger der auszutauschenden Daten über die geeigneten Lesegeräte verfügen, ist viel größer als bei vielleicht besonders eleganten, vielleicht besonders effektiven, aber dennoch selten anzutreffenden Speicherlösungen.

Mit der weiten Verbreitung des Medientyps (CD, Diskette, ...) ist schon ein wesentliches Kriterium genannt, welches bei einer Empfehlung von Speichermedien angewendet werden muss. Wichtig ist dabei eine große Verbreitung sowohl der Medien als auch der Schreib- und Lesegeräte, nur dies macht wahrscheinlich, dass auch in einigen Jahren noch beides erreichbar ist. Es lassen sich noch weitere Kriterien ermitteln:

- Wichtig ist selbstverständlich, dass die Medien eine große Lebensdauer haben.
- Die Medien sollten robust sein und
- ihre Lagerung sollte keine kaum zu erfüllenden Anforderungen stellen
- Die Medien sollten zudem über eine angemessene Kapazität verfügen, um Dateien in Gänze (d.h. ohne Komprimierung) aufzunehmen.
- Die Bedienung der Schreib- und Lesegeräte sollte leicht erlernbar sein.
- Die einzelnen Dateien und Daten sollten auf den Medien schnell erreichbar sein.
- Die Preise der Medien sollten möglichst niedrig sein. ...

Bevor im folgenden auf die einzelnen Arten von Speichermedien eingegangen wird, soll hier noch eine allgemeine Anmerkung festgehalten werden: Die technische Entwicklung bedingt, dass es „das eine“, „das ultimative“, „das endgültige“ Speichermedium nicht gibt. Eines ersetzt das andere. Im Bereich der Tonspeicherung beispielsweise wandelten sich die Trägermedien von Schellack- zu Vinylplatten, von Magnetbandkassetten zu CDs. Die mit den Medien verbundenen Geräte wandelten sich vom Grammophon über Schallplattenspieler, Kassettenrekorder zum CD-Player. Jeder Wechsel darin, welches Medium gerade handelsüblich war, zwang die Liebhaber bestimmter alter Musik, spätestens dann, wenn sein Abspielgerät unreparierbar defekt war, erneut Geld für die Möglichkeit, seine Musik zu hören, auszugeben. Ebenso ist es im Bereich digitaler Daten. Der Drang und Zwang der Industrie, Geld zu verdienen, wird auf absehbare Zeit stets neue Speichermedien mit neuen Lese- und Abspielgeräten hervorbringen. Wer das „endgültige“ Speichermedium findet, der verdient nur ein einziges Mal.

Alle gegenwärtigen Medien zerfallen durch Umwelteinflüsse. Egal, welches Medium verwendet wird, nach einer bestimmten Zeit wird ein Kopieren auf andere Medien gleicher Art (Refreshing oder Auffrischen) oder auf ein Medium anderer Art (Migration oder Datenwanderung) notwendig. Wie lange welches Medium dem natürlichen Verfall trotzen kann, ist notwendigerweise vage. Niemand kann es mit Gewissheit sagen. Die Hersteller sprechen von großen und sehr großen Zeitspannen, im Falle von CDs können das bis zu 200 Jahre sein. Doch was bedeuten diese Angaben?

Die Angaben beruhen nicht auf Erfahrungen, können es nicht, denn die Medien gibt es dafür noch nicht lang genug! Um Aussagen über die Haltbarkeit oder Lebensdauer ihrer Medien machen zu können, unterziehen die Hersteller ihre Medien so genannten Alterungstests. In diesen Tests werden die Medien über eine bestimmte Zeit hinweg hohen Temperaturen und hoher Luftfeuchtigkeit ausgesetzt. Anschließend wird ermittelt, wie groß die Veränderungen sind und wie viele der getesteten Medien noch uneingeschränkt lesbar sind. Auf dieser Basis werden Überschlagsrechnungen angestellt. Für solche Tests wurde erst vor kurzer Zeit, im Falle von beschreibbaren CDs erst im Jahre 2002, eine Norm (für CDs: ISO 18921:2002) geschaffen. CDs aber gibt es, seit sie 1980 von Philipps und Sony eingeführt wurden. Und es gibt sie seit damals mit Versprechen, dass sie eine Lebensdauer von 10, 50 oder 100 Jahren haben. Auch heute ist kein Hersteller daran gebunden, sich an die Norm der ISO zu halten, d.h. seine CDs genau diesen Testbedingungen auszusetzen und den Zerfallsgrad, ab welchem er eine CD als nicht mehr verwendbar bezeichnet, als vorgegeben zu akzeptieren. In der Praxis beruhen die Angaben der Hersteller über die Lebensdauer ihrer Speichermedien auf hauseigenen Tests unter hauseigenen Bedingungen mit einer Auswertung nach hauseigenen Kriterien! Doch selbst wenn alle sich an die genannte Norm hielten, bleibt zu fragen, wie viel ein Test, der sich auf Luftfeuchtigkeit und Temperatur beschränkt, wirklich aussagt. Andere Dinge spielen ebenfalls eine Rolle. Beispielsweise, wie oft ein Medium gelesen wird, wie

staubfrei die Umgebung ist, wie viel Schwefel oder andere chemische Bestandteile die Umgebungsluft hat, wie viel ultravioletter Strahlung das Medium in einer bestimmten Zeitspanne ausgesetzt ist, Es ist deshalb angebracht, die Alltagstauglichkeit dieser Tests zu hinterfragen und den Angaben der Hersteller nicht ohne weiteres zu vertrauen.

Sicher ist jedoch, dass die Lebensdauer von Speichermedien, jener Zeitraum, innerhalb dessen die Daten unversehrt aus dem Medium ausgelesen werden können, von der Art der Lagerung und vom Umgang mit dem Medium abhängt.

Disketten, Magnetbänder, Festplatten, CDs und DVDs sollen im Folgenden dahingehend betrachtet werden, ob sie als erster Baustein für eine Langzeitarchivierung (nicht für übermorgen, aber wenigstens für morgen) geeignet sind. Die einzelnen Medien werden mit den oben genannten Kriterien - Verbreitungsgrad, Robustheit, Kapazität, Zugriffsgeschwindigkeit, Preis - konfrontiert. Anschließend folgen zu den einzelnen Medientypen Ratschläge, wie die Lebensdauer möglichst lange aufrechterhalten werden kann.



Information, für lange Zeit gespeichert (Hieroglyphen aus Luxor)

Medien im Einzelnen

Diskette

Beschreibung:

Disketten sind ein seit vielen Jahren genutztes Medium. Sie haben allerdings mehrfach ihre Größe geändert, von 8 Zoll Kantenlänge über 5,25 Zoll zu den heute noch anzutreffenden 3,5 Zoll Disketten. In einer Plastikhülle ist eine Scheibe untergebracht, die mit einer magnetisierbaren Schicht versehen ist. In dieser Schicht werden die Daten durch Ausrichtung von Magnetfeldern gespeichert. Obwohl die Disketten kleiner wurden, erhöhte sich die Speicherkapazität, indem die Daten dichter nebeneinander platziert, d.h. die Magnetfelder kleiner, wurden. Ein mechanischer Schreib- und Lesearm tastet die Magnetfelder ab. Er berührt dabei die Oberfläche der dünnen Scheibe. Die Scheibe dreht sich mit ungefähr 300 Umdrehungen pro Minute in ihrem Gehäuse. An einer bestimmten Stelle (Sektor 0) ist das Inhaltsverzeichnis der Diskette gespeichert.

Verbreitung:

Nach wie vor sind Disketten weit verbreitet. Sie dienen dem Austausch kleiner Datenmengen (im Allgemeinen nicht mehr als 1,44 MB) zwischen Computern, die nicht über ein Netzwerk miteinander verbunden sind. Doch die Zeit weit verbreiteter Nutzung von Disketten geht ihrem Ende entgegen, einerseits gibt es insbesondere seit der Verbreitung digitaler Fotografie, immer häufiger Dateien, die nicht in Gänze von einer einzelnen Diskette aufgenommen werden können, und zum anderen ersetzen so genannte USB-Sticks, deren Speichervolumen wesentlich höher ist (variabel, gegenwärtig bis zu 1 GB) die Disketten als handliches, leicht beschreibbares Medium für den Datentransfer. Noch vor einem Jahr stand es außer Frage, dass ein neu erworbener Computer über ein Diskettenlaufwerk verfügte. Mittlerweile ist diese Ausstattung zwar noch üblich, jedoch nicht mehr per se zu erwarten.

Handhabung / Lagerung:

Auch wenn Disketten für die Langzeiterhaltung nicht empfohlen werden können, seien hier Empfehlungen genannt, wie Disketten möglichst lange lesbar bleiben:

- licht- und staubgeschützte Umgebung
- sichere Entfernung von magnetischen Feldern (Lautsprecher, älterer Monitor, ...)
- ca. 20 Grad Raumtemperatur
- 40 bis 45 Prozent Luftfeuchtigkeit
- berühren Sie die Scheibe, welche die Daten enthält, nicht mit den Fingern

- beschreiben Sie Disketten nicht mit harten Stiften durch kräftiges Aufdrücken, Sie könnten die Datenträgerscheibe zerstören

Unter diesen Bedingungen gehen manche Hersteller von einer Lebensdauer von 10 Jahren aus. Mit jedem Schreib- oder Lesezugriff verringert sich die Lebensdauer. Bei häufiger Benutzung sprechen die Hersteller von einer Lebenszeit von (geschätzt) 5 Jahren. Um sicher zu gehen, sollten Sie spätestens nach einem Jahr ihre Daten auf einen neuen Datenträger kopieren.

Abwägung:

Die Geschichte der Disketten ist eine wechselnder Außenmaße und schnell vom Markt verschwundener Lese- und Schreibgeräte. Disketten sind, schon allein, weil sie in absehbarer Zeit kaum noch gebräuchlich sein werden, eindeutig für die Langzeiterhaltung ungeeignet. Wer wichtige Daten auf Disketten gespeichert hat, der sollte diese möglichst bald auf ein anderes Speichermedium übertragen.

ZIP-Diskette

Beschreibung:

Für die so genannten ZIP-Disketten gilt im wesentlichen das, was schon zu Disketten im allgemeinen gesagt wurde. Sie waren dafür gedacht, das Problem der geringen Kapazitätsgrenze von Disketten zu lösen, indem die mögliche Speichermenge wesentlich erhöht wurde. Kurze Zeit danach wurde allerdings die beschreibbare CD, deren Speicherkapazität größer als die einer damaligen ZIP-Diskette war, auf den Markt gebracht. Einer weiten Verbreitung von ZIP-Disketten stand auch entgegen, dass die einzelnen Medien relativ teuer blieben. Dies wiederum war der Tatsache geschuldet, dass die Lizenzen und Patente für die ZIP-Technologie im Wesentlichen bei nur einer Firma (Iomega) lagen, was Konkurrenz ausschloss. Die Empfehlungen für die Lagerung entsprechen denen von Disketten (s.o.)

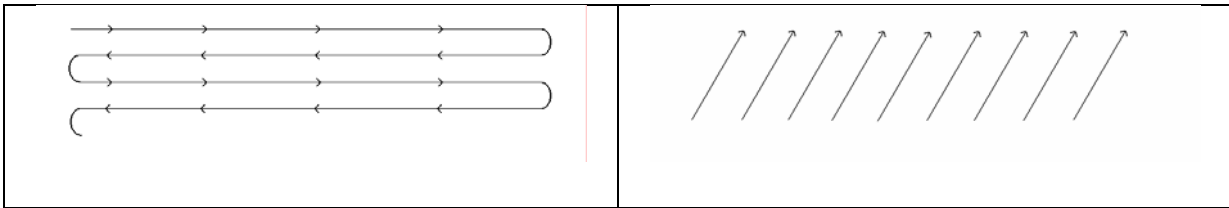
Abwägung:

ZIP-Disketten sind, wie andere Disketten auch, ungeeignet für die Langzeitarchivierung. Wer ZIP-Disketten verwendet, setzt sich zudem der Produktpolitik einer einzigen Firma aus, die jederzeit die Produktion entsprechender Medien oder der nötigen Lese- und Schreibgeräte einstellen kann.

Magnetband

Beschreibung:

Während Disketten die Daten auf einer magnetisierbaren Scheibe speichern, befindet sich die magnetisierbare Schicht von so genannten Magnetbändern auf einem Band, welches auf zwei Spulen aufgewickelt wird. Grundsätzlich gibt es verschiedene Verfahren, die Daten auf den Bändern anzuordnen, entweder



hintereinander in einer langen Schleife oder schräg nebeneinander. Die Art der Anordnung der Daten hat vor allem auf die Lesegeschwindigkeit und die Häufigkeit des Vor- und Zurückspulens Einfluss. Belastungen, denen die Bänder ausgesetzt sind, entstehen durch den Spulungsvorgang, aber auch dadurch, dass der Lesekopf auf dem Band aufsetzt. Die Zeitspanne, innerhalb derer gespeicherte Daten ohne Verluste zurückgelesen werden können, hängt deshalb wesentlich von der Häufigkeit des Zugriffs ab.

Magnetbänder werden schon seit den 1950er Jahren eingesetzt - seit Mitte der 1980er Jahre nicht mehr vornehmlich in offenen Spulen, sondern in so genannten Cartridges. Es gibt Geräte, welche automatisch in vorgegebenen Zeitabständen Daten von alten auf neue Magnetbänder kopieren. Auf diesem Weg lässt sich die Langzeiterhaltung von Daten wenigstens partiell automatisieren.

Verbreitung:

Die Sicherung von Daten, im Sinne einer Sicherheitskopie, geschieht vielerorts mittels Magnetbändern. Vor allem bei größeren Datenmengen kommen solche Bänder zum Einsatz. Bandroboter werden dabei häufig von mittleren und größeren Institutionen eingesetzt.

Handhabung / Lagerung:

- eine möglichst staubfreie Umgebung
- aggressive Gase und Chemikalien schaden den Magnetbändern
- erst nach einer Akklimatisierungsphase sollten Bänder, die einer anderen klimatischen Umgebung entnommen wurden, in ein Laufwerk eingeführt werden
- die Luftfeuchtigkeit sollte unter 60 Prozent und die Temperatur um 18-20° betragen
- die Bänder immer aufrecht stellen und niemals mehrere übereinander legen

- die Gleichmäßigkeit der Spulung wird durch Herunterfallen der Bänder leicht beeinträchtigt, was zu Lesefehlern und auch zu Bandriss führen kann
- Berührungen der Bandoberfläche führen leicht zu Datenverlust
- Werden Magnetbänder über längere Zeiten nicht umgespult, kann es zu einem „durchdrücken“ kommen. Die Magnetisierung einer Bandstelle erscheint auf einer anderen, die, beispielsweise in einer Spule innerhalb einer Kassette, direkt auf dieser aufliegt
- Bewahren Sie Magnetbandkassetten nicht in Papier oder Pappschachteln auf. Spezielle Plastikgehäuse weisen Staubpartikel wesentlich besser ab. Solche Aufbewahrungsschachteln müssen dicht schließen und die Kassetten fest halten.
- in Abhängigkeit von der Häufigkeit des Lesens und Beschreibens der Medien ist ein Austausch der Medien unerlässlich. Spätestens einmal im Jahr sollte zur Sicherheit ein solcher Austausch vorgenommen werden, auch wenn das Medium kaum genutzt wurde.

Abwägung:

Gegenwärtig wird die Kapazität der Speicherbänder ständig erhöht. Die Daten werden dichter auf die Medien geschrieben. Oft erfordert dieses Veränderungen nicht nur der Software zum Lesen und Beschreiben der Medien, sondern auch Veränderungen der Hardware (Medien sowie Lese-/Schreibgeräte). Die Sicherung von Daten auf Magnetbändern unter Verwendung von Bandautomaten ist die gegenwärtig übliche Methode der längerfristigen Datenerhaltung in mittleren und größeren Institutionen und Firmen.

Die Möglichkeit des automatisierten Kopierens erscheint zunächst als großer Vorteil – die Problematik, dass Dateien in neue Formate übertragen werden müssen, wird dabei allerdings nicht berücksichtigt. Zu den Nachteilen von Magnetbändern gehört die zum Teil recht lange Zugriffszeit und die Abhängigkeit der Nutzbarkeitsdauer von der Häufigkeit des Zugriffs. Der recht schnelle technologische Wandel in der Magnetbandtechnologie, der dazu führt, dass für Medien, die nur wenige Jahre alt sind, keine Lese- und Schreibgeräte mehr zur Verfügung stehen, ist in Hinblick auf die Langzeiterhaltung gleichfalls als großer Nachteil anzusehen. Für jene, die häufiger auf Daten zurückgreifen müssen, die sie für die Langzeiterhaltung gespeichert hatten, und für jene, die nicht über einen Bandautomaten verfügen, sind Magnetbänder wenig geeignet, da der Arbeitsaufwand für die dann häufig nötigen Kopiervorgänge groß ist, die Zugriffszeiten lang sind und der Erwerb neuer Datenträger in kurzen Intervallen teuer ist.

Festplatte

Beschreibung:

Ähnlich wie Disketten sind Festplatten aufgebaut. Mehrere Schichten von Scheiben, die mit einer magnetisierbaren Schicht versehen sind, liegen dabei übereinander. Im Unterschied zu Disketten ist der Schreib- und Lesekopf integriert und das System aus Scheiben und Köpfen weitgehend geschlossen, so daß kein Staub eindringen kann. Die feste Bauweise erlaubt es, die Schreib-Leseköpfe in einem sehr kleinen Abstand an den Scheiben entlang zu führen. Die einzelnen Scheiben werden dabei nicht berührt. Durch dieses reibungs- und berührungsfreie Verfahren lassen sich wesentlich höhere Zugriffszeiten realisieren als bei Disketten oder Magnetbändern möglich. Festplatten drehen sich mit mehreren tausend Umdrehungen pro Minute recht schnell. Entsprechend hoch ist die mechanische Belastung des Materials. Die Bauweise bedingt jedoch eine hohe Empfindlichkeit gegenüber Stößen. Festplatten wurden, ähnlich wie Disketten, im Laufe der Jahre immer kompakter, können bei immer kleineren Ausmaßen immer größere Datenmengen aufnehmen.

Festplatten gibt es als fest in ein Rechnersystem integrierte Platten oder in Form von Wechselplatten, bei denen die gesamte Mechanik und Steuerelektronik in einem Gehäuse aus dem Rechnersystem genommen und an anderen Orten (oder zu anderen Zeiten) wieder eingesetzt werden kann.

Handhabung / Lagerung:

- Stöße und Schläge sind kritisch
- Magnetfelder sind zu vermeiden
- Bei Wechselfestplatten ist eine Akklimatisationsphase an die Luft- und Raumumgebung einzuhalten.

Abwägung:

Festplatten verbinden eine hohe Speicherkapazität mit einer schnellen Zugriffszeit. Als Speichermedium für lange Zeiträume eignen sie sich nur bedingt, da sie entweder in ein Rechnersystem integriert oder, wenn nicht, recht stoßanfällig sind. Seit die Preise für Festplatten rapide gesunken sind, treten sie in immer mehr Institutionen in Konkurrenz zu Magnetbändern als Mittel zur Datensicherung. Dort, wo auf (Langzeit-) archivierte Daten schnell und häufig zugegriffen werden muss, sind die Platten den Bändern überlegen. Wo dies nicht der Fall ist, sind – vor allem preislich – noch immer Magnetbänder die günstigere Alternative, denn sie lassen sich leichter aus den Lese-/Schreibgeräten entnehmen und an sichere Orte verbringen.

CD

Beschreibung:

CDs gibt es seit Anfang der 1980er Jahre. Die ursprünglich vor allem von Sony und Philipps entwickelte Technologie wird heute von vielen Herstellern verwendet, dies gilt für Lese- und Schreibgeräte ebenso wie für die entsprechenden Medien.

Unter den verschiedenen Arten von beschreibbaren CDs kommen vor allem die CD-R, die nur einmal beschrieben werden kann, und die CD-RW, die mehrfach beschrieben werden kann, für die langfristige Sicherung von Daten in Frage. Prinzipiell ist jede CD nach dem gleichen Schema aufgebaut: Auf der Unterseite befindet sich eine Schutzschicht, die durchsichtig und im allgemeinen aus Polycarbonat ist. Darüber liegt eine dünne Schicht, die Daten-Trägerschicht, für die verschiedene Hersteller verschiedene Materialien verwenden. Auf dieser Daten-Trägerschicht liegt eine Lackschicht auf, welche das Licht des Lasers reflektiert. Die nächste Schicht dient wiederum dem Schutz, sie ist oftmals bedruckt, und schließt die CD nach oben ab.



Das Licht des Lasers durchdringt von unten das Material. Wenn Daten geschrieben werden, erzeugt es durch Hitzeeinwirkung eine Veränderung der Oberfläche in der Datenschicht. Wenn Daten gelesen werden, wird der Laserstrahl durch solche Unebenheiten in der Datenschicht zerstreut oder er erreicht die Lackschicht und wird reflektiert. Aus den Erhö-

hungen und Vertiefungen, umgesetzt in Lichtsignale eines Laserstrahls, werden die Daten rekonstruiert. Unterschiede zwischen CDs ergeben sich aus der Verwendung verschiedener Materialien für die Reflektor- (gebräuchlich: Gold, Silber, Aluminium) und für die Datenträger-Schicht (gebräuchlich: Cyanin und Azon). Die jeweilige Kombination der für diese zwei Schichten verwendeten Materialien bestimmt die Farbigkeit der Unterseite. Je nach Hersteller und Fabrikat kommen unterschiedliche Werkstoffe in unterschiedlichen Mischungsverhältnissen zum Einsatz. Für die mehrmals beschreibbaren CD-RWs werden andere Stoffe in der Datenträgerschicht verwendet. Audio-CDs haben eine andere logische Aufteilung als Daten-CDs. Industriell gepresste CDs sind, anders als gebrannte CDs, nicht mit Erhöhungen und Vertiefungen, sondern mit gezielt lichtundurchlässig gemachten Stellen versehen. Das Prinzip des Aufbaus ist jedoch in allen Fällen ähnlich. Das Arbeiten mit Licht ermöglicht einen größeren Abstand zwischen Medium und Lesekopf, als dies beispielsweise bei Festplatten möglich ist. Die Daten werden stets hintereinander, in Form einer Spirale (Gesamtlänge ca. 6 km) auf einer CD abgelegt.

Weil die Hersteller die Rezepturen der verwendeten Materialien für die Datenträgerschicht laufend variieren und weil manche CD-Lesegeräte besser CDs der einen, andere Geräte aber besser CDs einer anderen Rezeptur lesen können, ist es schwer zu sagen, ob golden schimmernde oder silbern leuchtende CDs besser sind.

Trotz prinzipiell gleichen Aufbaus sind CD-RWs etwas komplizierter gestaltet. Unter- und oberhalb der Datenträgerschicht verfügen sie über eine dünne Hilfsschicht, die Zusammensetzung des Daten-Trägermaterials (gebräuchlich sind Mischungen aus: Silber, Indium, Antimon, ...) ist eine andere und es werden statt Erhöhungen und Vertiefungen vom Laserstrahl kleine lokale Veränderungen in der Kristallstruktur der Trägerschicht erzeugt.

Das Fassungsvermögen von CDs liegt bei 650 oder 700 MB. Das entspricht vielen Millionen Erhöhungen und Vertiefungen, die auf einer Scheibe mit nur 12 cm Durchmesser Platz finden müssen. Die einzelnen Vertiefungen sind nur 0,9 Millionstel Meter lang. Die Schreib- und Lesevorrichtung muss entsprechend genau arbeiten. Diese große Datendichte bedingt allerdings auch eine große Anfälligkeit für Kratzer, Verbiegungen, Schmutz oder Fingerabdrücke. Alles, was den Lauf des Lichtstrahles ungezielt verändert, wird entweder nicht oder falsch gedeutet. Im Allgemeinen sind deshalb CD-Laufwerke mit Fehlerkorrekturen ausgestattet, die es erlauben, einzelne fehlende oder falsch interpretierte Informationen aus mit ihnen gespeicherten Prüfsummen zu rekonstruieren. Doch schon bei einem etwas tieferen Kratzer oder einem fest haftenden Fingerabdruck reicht dieses oft nicht aus.

Trotz Fehlerkorrekturprogrammen ermöglicht die CD-Technologie ein – im Vergleich zum Magnetband – schnelles Zugreifen auf gespeicherte Daten. CDs sind zudem vergleichsweise preisgünstig.

Verbreitung:

CDs, in welcher Form auch immer, sind sehr weit verbreitet. Es gibt die entsprechenden Lese- und Schreibgeräte für alle gängigen Betriebssysteme.

Robustheit:

Angesichts der Genauigkeit, mit der die CD-Technologie arbeitet, scheinen CDs erstaunlich robust zu sein. Oftmals bleibt es aber beim Schein. Im Allgemeinen gilt, dass gepresste CDs wesentlich robuster sind als gebrannte und dass CD-Rs wesentlich robuster sind als CD-RWs. Letztere sind oft schon nach kurzen Gebrauchszeiten nicht mehr uneingeschränkt lesbar. Da andererseits die entsprechende technische Ausstattung zum Erstellen gepresster CDs in den meisten Fällen den finanziellen Rahmen sprengt, beziehen sich die folgenden Hinweise zur Handhabung und Lagerung ausschließlich auf selbstgebrannte CDs.

Handhabung / Lagerung:

Mancher der folgenden Hinweise ist auch auf Verpackungen von Datenträgern zu finden - andere nicht.

- Direktes Sonnenlicht kann zu chemischen Veränderungen führen und sollte vermieden werden
- Starke Hitze und hohe Luftfeuchtigkeit können zu einer unterschiedlich starken Ausdehnung der einzelnen Schichten und damit zu einer Krümmung der Oberfläche führen. Chemische Reaktionen können in Gang gesetzt werden. Vom äußeren Rand her kann Feuchtigkeit zwischen die einzelnen Schichten dringen und auf diese Weise zu Unebenheiten führen. Die besten Lagerbedingungen bestehen bei 20% bis 50% Luftfeuchtigkeit und einer Temperatur um oder unter 20 Grad.
- Schnelle Veränderungen von Temperatur oder Luftfeuchtigkeit sind zu vermeiden, sie lassen die Werkstoffe leicht spröde werden
- Lagern Sie CDs in senkrechter Position in den dafür vorgesehenen Plastikgehäusen, vermeiden sie in jedem Falle CDs aufeinander zu legen. Durch Druck können Deformationen entstehen.
- Fingerabdrücke, vor allem auf der unteren Seite, führen zu Lesefehlern
- Kugelschreiber oder ähnliche Stifte erzeugen beim Beschreiben Druckstellen, die als Vertiefungen in der Reflektorschicht zu einer Umlenkung des Laserlichts führen. Es gibt lösungsmittelfreie Stifte auf Wasserbasis, verwenden Sie ausschließlich diese.
- Etiketten aufzukleben ist nicht ratsam, denn sie sind selten wirklich zentriert, was zu Ungleichgewichten bei der Rotation führt und damit das Lese- und Schreibgerät belastet, zudem bleiben die verwendeten Klebstoffe nicht ohne Einfluss auf die obere Schutzschicht. Es kann zu chemischen Reaktionen kommen, bei denen Stoffe in die Reflektorschicht eindringen und diese verändern. Schließlich können sich die Etiketten lösen und so zu größeren Schäden am Lese- Schreibgerät führen.
- Beschreiben Sie nur den durchsichtigen inneren Ring einer CD. Prüfen Sie, ob die Fläche für Ihre Bedürfnisse ausreicht.
- Packen Sie CDs, welche Sie beschreiben wollen, erst kurz vor dem Beschreiben aus um die Wirkung von Umwelteinflüssen zu minimieren. Prüfen Sie vor dem Beschreiben die Oberfläche (besonders die Unterseite). Oft lässt sich ein Fehler (z.B. Kratzer) mit bloßem Auge erkennen.
- Sofort nach Benutzung sollten CDs aus dem verwendeten Laufwerk entnommen werden (Computer und Laufwerk bieten andere klimatische Bedingungen als die Plastikgehäuse).
- Prüfen Sie nach dem Schreibvorgang die Lesbarkeit der CD-R. Verwenden Sie dazu am besten ein anderes Laufwerk als jenes, mit dem Sie die Daten geschrieben haben. Öffnen Sie wenigstens einige der geschriebenen Dateien.

- Der Schreibvorgang kann zu Fehlern führen. Im Allgemeinen ist eine so genannte vierfache Geschwindigkeit angemessen.
- Verwenden Sie keine mehrfach beschreibbaren Medien (CD-RW). Diese sind wesentlich anfälliger für Datenverluste als die nur einmal beschreibbaren CD-Rs.
- Ob Daten gelesen werden können, hängt auch vom Zusammenspiel zwischen Lese-/Schreibgerät und CD-R ab. Oft sind jene CD-R-Typen, welche die Hersteller der CD-Schreibgeräte empfehlen, besonders für das entsprechende Gerät geeignet. Sollten keine Empfehlungen gegeben worden sein, dann testen Sie verschiedene CD-Typen, indem Sie mit Ihrem Schreibgerät Daten darauf schreiben und anschließend mit verschiedenen Lesegeräten versuchen, diese Daten zu lesen.
- Wenn Ihr CD-Lesegerät die Daten aus einer CD-R nicht lesen kann, dann sind nicht notwendigerweise die Daten verloren. Versuchen Sie es in einem anderen, möglicherweise neueren, Gerät, das eventuell über eine bessere Fehlerkorrektur verfügt und brennen Sie anschließend – falls dieses Gerät die Daten lesen konnte – eine neue CD-R.
- Reinigungsmittel sind daraufhin auszuwählen, ob sie mit den Materialien der CDs chemisch reagieren. Harte Tücher oder grobe Pinsel führen zu Kratzern, die groß genug sind, um Daten zu zerstören.
- Reinigen Sie eine CD stets von innen nach außen oder umgekehrt, jedoch nie in konzentrischer Form. Sie folgen sonst der Anordnung der Daten, weshalb eventuell erzeugte Kratzer gleich ganze Blöcke von Daten zerstören würden
- Sauberpusten transportiert stets kleine, bisweilen kaum sichtbare, Mengen Spucke, die sich auf den CDs ablagert und das Lesen beeinträchtigt.
- Lesefehler resultieren bisweilen auch aus Verschmutzungen auf der Linse des Lasers. Bevor Sie die Daten auf einer CD verloren geben, reinigen Sie die Linse (Reinigungssets im Handel) oder lassen Sie dies von Fachleuten erledigen.
- Kopieren Sie Ihre Daten spätestens nach 2 bis 3 Jahren auf eine neue CD oder ein anderes Speichermedium.

Abwägung:

Die allgemeine Erfahrung zeigt, dass selbstgebrannte Daten-CDs weniger langlebig sind, als die Hersteller der Rohlinge versprechen. Dennoch halten sie einem Vergleich mit anderen Medien stand. Wie andere hier besprochene Medien sind auch CDs von Umwelteinflüssen abhängig. Sie sind vergleichsweise preiswert, sind transportabel und weit verbreitet. Für die CDs zeichnet sich dennoch ein technologischer Wandel ab. Die neuesten Computer werden nicht mehr wie selbstverständlich mit einem CD-Schreib-/Lesegerät ausgestattet. Stattdessen verfügen sie über DVD-Schreib-/Lesegeräte, welche auch CD-Rs lesen und beschreiben können. Da CDs mit einer geringeren Geschwindigkeit arbeiten und die Daten weniger dicht auf die Scheibe packen als DVDs, bedeutet dies, dass die CD-Technologie für die DVD-Lese-/Schreibgeräte wie eine Bremse funktioniert - um CDs lesen

und schreiben zu können, müssen die Geräte so gebaut sein, dass sie auch langsamer als für DVDs gebräuchlich arbeiten können.

Es steht zu erwarten, dass sich schon mittelfristig schnellere Lese- und Schreibgeschwindigkeiten für DVDs durchsetzen. Die Lese- und Schreibgeräte werden daraufhin optimiert werden. Die Bremse „CD“ steht dann zur Disposition. Es erscheint deshalb bereits mittelfristig möglich, dass die Möglichkeit, CDs zu lesen, nicht mehr überall gegeben ist. Wer gegenwärtig ein DVD-Lese-/Schreib-Laufwerk kauft, sollte darauf achten, dass es in der Lage ist, mit möglichst vielen CD-Formaten umzugehen.

DVD

Die DVD, deren Spezifikationen 1995 festgeschrieben wurde, ähnelt in vielem der CD, das geht über den gemeinsamen Durchmesser und die gemeinsame kreisrunde Form hinaus. Wie CDs sind DVDs in Schichten aufgebaut. Während alle Arten von CDs stets nur eine Daten-Träger-Schicht haben, können es bei einer DVD bis zu vier Schichten sein, in denen mittels Laser Vertiefungen angebracht werden. Der Laser arbeitet etwas anders als jener, der bei CDs eingesetzt wird. Durch Licht anderer Wellenlänge lassen sich mehr Informationen auf dem gleichen Raum unterbringen. Die Vertiefungen sind kleiner und liegen näher beieinander, als es bei CDs der Fall ist. DVD-Laufwerke drehen zudem mit einer höheren Geschwindigkeit.

Durch die höhere Datendichte und auch, weil manche DVD-Hersteller aus Kostensenkungserwägungen auf Schutzschichten verzichten, sind DVDs kratzempfindlicher und reagieren heftiger auf UV-Licht als CDs (Stiftung Warentest: <http://www.warentest.de>). Prinzipiell gibt es unter den Herstellerfirmen zwei Lager, jene, die DVD-R (und DVD-RW) favorisieren, und jene, die DVD+R (sowie DVD+RW) zum allgemeinen Standard erheben möchten. Auch weil sich hier noch kein internationaler Standard durchgesetzt hat, bleibt die DVD im Hinblick auf längerfristige Datenerhaltung ein fragwürdiges Medium, denn wenn Sie sich für den „falschen“ Standard entscheiden, d.h. DVDs in jenem Format brennen, welches sich nicht durchsetzt, so werden Sie in einiger Zeit Schwierigkeiten bekommen, Ihre Daten zu lesen. Auch andere Probleme bleiben ungelöst. Die sogenannte DVD-9 – Form bereitet zum Beispiel vielen Abspielgeräten Probleme.

Weil die technische Entwicklung bei DVDs in einem sehr schnellen Fluss – und eigentlich schon überholt - ist, kann die DVD-Technologie weder für eine lang- noch für eine mittelfristige Datenarchivierung empfohlen werden. Sollten Sie dennoch DVDs als Medium für mittel- bis langfristiges Speichern verwenden wollen, dann sollten Sie in Handhabung und Lagerung so verfahren, wie es soeben im Abschnitt über CDs gesagt wurde.

Blu-ray Disc

Blu-ray Discs (oder Blu-ray, bisweilen auch als BD bezeichnet) sind das jüngste Mitglied in der Gruppe der optischen Datenträger. Als Speichermedium für so genanntes High-Definition Video und für Computerspiele geschaffen, haben Blu-ray die gleichen Ausmaße wie CD oder DVD. Wie bei CD und DVD gibt es auch bei Blu-ray beschreibbare (BD-R) und wiederbeschreibbare Datenträger (BD-RE). Letztere kamen 2006 auf den Markt. Der Name dieser Speichermedien rührt vom Laser, mit welchem die Scheiben gelesen und geschrieben werden. Dieser Laser verwendet Licht mit einer Wellenlänge im blauen Bereich. Blu-ray wurden von der Blu-ray Disc Association, einer Gruppe großer Firmen bereits in 2002 entwickelt. Bisher konnten Sie nur langsam einen erwähnenswerten Marktanteil erobern.

Auch wenn Blu-ray Discs dadurch beeindrucken, dass sie wesentlich mehr als beispielsweise eine DVD zu speichern vermögen sind sie für die Langzeitarchivierung ungeeignet. Binnen weniger Jahre dürften Blu-ray durch günstigere Medien ersetzt werden, weshalb die entsprechenden Lese- und Schreibgeräte bald nicht mehr allgemein zur Verfügung stehen dürften. Insbesondere die gesunkenen Preise für immer größere Festplatten stehen einer nennenswerten Marktpräsenz der Blu-ray entgegen. Das Problem der Dichte, mit welcher Daten auf eine Scheibe geschrieben werden, zeigt sich hier erneut und in verschärfter Weise.

Blu-ray sind, alles in allem, für die Langzeitarchivierung ungeeignet. Falls sie dennoch für diesen Zweck verwendet werden sollten, dann sind die Empfehlungen für die Aufbewahrung von CDs auch hier anzuwenden.

USB - Sticks

Allein ihrer aktuell sehr weiten Verbreitung wegen sind USB-Sticks an dieser Stelle eine Erwähnung wert. Es gibt sie seit 2000 auf dem Markt. USB-Sticks sind vergleichsweise kostengünstig und es gibt unzählige Varianten die immer größere Datenmengen aufnehmen können. Die Zahl der Schreib- und Lesevorgänge ist begrenzt und es gibt keine Gewähr für die Beständigkeit der Daten. Angaben der Hersteller über die Beständigkeit von USB-Sticks können nicht vorbehaltlos übernommen werden und beziehen sich ohnedies meist nur auf eine Anzahl von Schreib- und Lesezyklen.

Ein anderes, in der Praxis nicht irrelevantes Faktum, ist, dass USB-Sticks sehr leicht verloren gehen, verlegt oder irgendwo vergessen werden. USB-Sticks sind für den Austausch von Daten sinnvoll, für die Langzeitarchivierung sind sie ungeeignet.

Zusammenfassung: Speichermedien

Auf den vorherigen Seiten wurden einige Arten von Speichermedien betrachtet. Es gibt noch weitere, die jedoch oft sehr kostenaufwändig sind oder deren Verbreitung sehr gering ist. Die Betrachtung hat vor allem eines gezeigt:

Aus der Sicht der Langzeiterhaltung digitaler Daten ist die Wahl des Speichermediums nicht von entscheidender Bedeutung!

Das eine Medium, jenes, dem wir unsere Daten anvertrauen und ohne weiteres Zutun nach vielleicht 100 Jahren wieder abgewinnen können, dieses Medium gibt es nicht. Alle gegenwärtig zur Verfügung stehenden Medien verlangen zwingend ein Kopieren der Daten auf neue Träger desselben oder eines anderen Medientyps nach einer bestimmten Zeitspanne.

Mit der Wahl des Medientyps, der für die Langzeiterhaltung eingesetzt werden soll, wird darüber entschieden, wie häufig solche Kopien erstellt werden müssen. Spätestens nach fünf Jahren müssen die Daten, die erhalten werden sollen, auf neue Datenträger übertragen werden, gleichgültig, welches Medium Sie verwenden. Daten auf Magnetbändern müssen häufiger auf neue Datenträger übertragen werden als Daten auf CDs. Die Entscheidung für einen Medientyp beeinflusst das Maß an Zeit und Arbeit, das für die Erhaltung der Daten aufgewendet werden muss. Hinzu kommen finanzielle Aspekte, wobei die Unterschiede jedoch nicht gravierend sind.

Die Betrachtung der verschiedenen Medientypen zeigt, dass fortwährende Wachsamkeit nötig ist. Die technische Entwicklung bringt ständig neue Medientypen oder Sub-Typen hervor. Alte Medientypen werden verdrängt und die notwendigen Lese- und Schreibgeräte verschwinden vom Markt. In diesem Falle kann es notwendig sein, schon lange bevor der Zeitpunkt für ein Überspielen auf ein neues Medium gekommen ist, eine Kopie der Daten auf einem Datenträger eines anderen Medientyps zu erstellen. Die Gefahr, davon überrascht zu werden, dass keine Lese-, Schreibgeräte mehr verfügbar sind, lässt sich dadurch verringern, dass von Anfang an die Daten mindestens zweifach gesichert werden, wobei wenigstens zwei verschiedene Medientypen Verwendung finden.

Mit den gegenwärtig verfügbaren Arten von Speichermedien ist dem Zwang zum Kopieren nicht zu entgehen. Die Intervalle zwischen Kopiervorgängen müssen beachtet werden. Um dies sicherzustellen, sollte unbedingt jedes für die längerfristige Archivierung vorgesehene Speichermedium mit dem Datum versehen werden, zu dem es mit Daten versehen wurde. Wenn möglich, sollte auch ein Protokoll erstellt werden, worin festgehalten wird, wann der nächste Kopiervorgang nötig ist. Schließlich sollte jemand

benannt werden, der die Sicherung überwacht und sicherstellt, dass zum gegebenen Zeitpunkt kopiert wird.

Ohne eine Festlegung von Verantwortlichkeiten und eine Benennung konkreter Arbeitsschritte, die zu festgelegten Zeitpunkten durchzuführen sind, lässt sich eine Langzeiterhaltung nicht gewährleisten!

Alles Aufheben – Das ist unmöglich

Jedem, der seinen Email - Eingang betrachtet, leuchtet sofort ein, dass es nicht sinnvoll ist, alles bewahren zu wollen – jedenfalls nicht für alle Ewigkeit. Auswahl tut not. Auch in der realen Welt wäre der Versuch alles für immer aufzubewahren erfolglos. In der digitalen Welt, in der problemlos Kopien von Kopien der Kopien einer Kopie hergestellt werden können ist die Auswahl dessen, was zu bewahren ist, von entscheidender Bedeutung. Der Punkt, an welchem die Übersicht über die Vielzahl mehr oder weniger identischer Kopien verloren geht, kommt unweigerlich. Zugleich ist es nicht immer legal, alles zu bewahren: Persönliche Daten Dritter, dem Urheberrecht unterliegende Dokumente, ... sind nur zwei Beispiele.

Unterziehen Sie die elektronischen Dokumente in ihrer Einrichtung einer eingehenden Prüfung und entscheiden Sie, welche davon nur von kurz- oder mittelfristigem Interesse sind. Entwickeln Sie auf dieser Grundlage Regeln nach welchen entschieden werden kann, welche Art von Information in welcher Form für lange Zeit bewahrt werden soll. Diese Regeln müssen sich selbstverständlich an den Zielsetzungen und den technischen Möglichkeiten ihrer Einrichtung orientieren. Machen Sie diese Regeln unter den Kollegen bekannt – damit alle sie befolgen können. Legen Sie fest, wann die gefundenen Regeln einer Prüfung auf Aktualität unterzogen werden müssen.

Entwerfen Sie eine Struktur für die Dateiablage. Wenn Sie hierfür ein einziges System verwenden, dann wird es später leichter, gesuchte Dateien wieder zu finden. Es ist weniger wichtig ob die Dateiablage für die Langzeitarchivierung zum Beispiel nach inhaltlichen Kriterien (etwa: Ein Ordner pro Thema) oder z.B. nach Zeitraum der Erstellung von Dateien (etwa: Ein Ordner für alle Dateien aus einem Monat), oder ... gegliedert ist. Wichtiger ist es, die gewählte Struktur einfach zu halten und sie nicht ständig zu verändern. Halten Sie auch diese Regeln schriftlich fest.

Kleine Schritte wie diese können wesentlich zur Langzeiterhaltung ihrer Daten beitragen. Entscheidend aber ist, wie bereits gesagt, anzuerkennen, dass Langzeitarchivierung wesentlich mehr ist, als die einmalige Wahl der richtigen Hard- oder Software. Sie ist vielmehr eine Frage der Festlegung von Verantwortlichkeiten, der Definition von Regeln, der Finanzierbarkeit und natürlich des Personal und der entsprechenden Fachkenntnis.

Eine Verlagerung der Aufgabe an eine Einrichtung außerhalb Ihrer Institution ist nur in Teilen möglich, denn die Aufgabe in regelmäßigen Abständen zu überprüfen, ob die eigenen Daten von einer solchen Einrichtung tatsächlich angemessen aufbewahrt werden, bleibt ebenso bei Ihnen wie auch die Abwägung, ob die aufgestellten Regeln einer Überarbeitung bedürfen.

Weitere Überlegungen und Empfehlungen

Bei jeder Änderung der Datenumgebung, dem Wechsel des Betriebssystems, der Einrichtung eines neuen Rechners oder der Verwendung eines neuen Programmes, sollten Sie sich fragen, ob jene Daten, die Sie für längere Zeit sichern wollen, noch uneingeschränkt verfügbar sein werden. Vielleicht müssen Sie handeln. Wenn beispielsweise der neue Rechner kein Laufwerk für die alten Datenträger hat oder wenn das Laufwerk des neuen Rechners die mit dem Schreibgerät des alten Rechners erstellten CDs nicht lesen kann oder wenn das neue Programm die alten Dateien nicht öffnen kann oder In jedem Fall sollten Sie bei jeder Änderung der Datenumgebung einen Test durchführen, denn solange der alte Rechner noch funktionstüchtig ist, kann er gegebenenfalls eingesetzt werden, um Ihre Daten aus den Medien zurückzugewinnen und sie in die neue Umgebung ohne Verluste zu transferieren.

Auch nach jedem Kopiervorgang sollten Sie einen Test vornehmen. Nicht jeder Schreibfehler wird vom Rechnersystem selbst erkannt. Versuchen Sie stets wenigstens einige der kopierten Dateien gleich nach dem Schreiben zu lesen. Wenn Sie die Möglichkeit haben, verwenden Sie dazu ein anderes als das zum Schreiben benutzte Schreib-/Lesegerät.

Speichern Sie prinzipiell Ihre für eine Langzeiterhaltung vorgesehenen Daten mehrfach! Speichern Sie auf verschiedenen Speichermedien! Lagern Sie die ausgewählten Medien an geographisch verschiedenen Orten unter den für die Medien jeweils besten Bedingungen. Die Mehrzahl der Orte hat den Vorteil, das aus Naturkatastrophen, Bränden oder auch Diebstählen resultierende Risiko des Datenverlustes zu reduzieren.

Versuchen Sie, „allgemein“ zu bleiben. Verwenden Sie allgemein verbreitete Software, ein weit verbreitetes Betriebssystem und einen Typ von Speichermedien, der oft anzutreffen ist. Speichern Sie Ihre Daten in einem möglichst allgemein verbreiteten und von möglichst vielen Programmen les- und interpretierbaren Format.

Berücksichtigen Sie schon beim Erstellen der Dateien den Langzeitaspekt, indem Sie Speicherformat und Speichermedium mit Bedacht wählen.

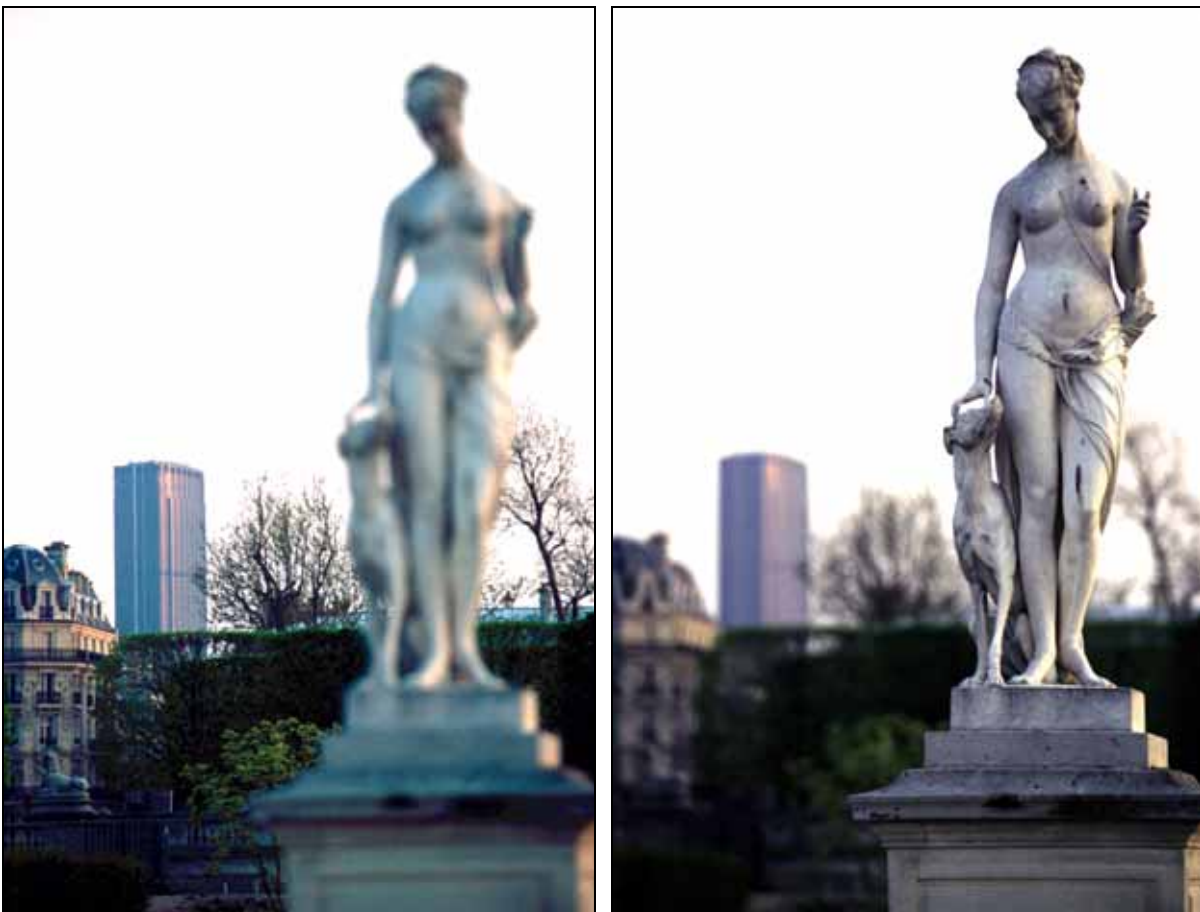
Nicht alle Daten müssen für eine lange Zeit aufgehoben werden. Prüfen Sie Ihre Datenbestände und legen Sie fest, welche Daten nur von kurzfristigem Interesse sind und welche langfristig aufbewahrt werden müssen.

Entsorgen Sie alte, aber noch funktionstüchtige Rechner und Programme nicht sofort, es kann passieren, dass Sie noch einmal auf diese zurückgreifen müssen.

Doch das wichtigste von allem:

Nehmen Sie Abschied von der Idee, Daten zu speichern, sei identisch damit, Daten für lange Zeit zur Verfügung zu haben. Sehen Sie die Langzeiterhaltung Ihrer Daten als eine eigenständige Aufgabe an, die nicht mit einer einmaligen Anstrengung zu bewältigen ist. Sehen Sie diese Aufgabe als einen Prozess, den Sie aktiv steuern müssen. Benennen Sie unbedingt Verantwortlichkeiten und erstellen Sie einen Plan.

Nur so geben Sie sich selbst und anderen die Chance, Ihre Daten auch noch in einigen Jahren nutzen zu können.



Vor der Zukunft steht die Vergangenheit

Anhang

Übersicht der Empfehlungen

Formate

Bilddateien	- TIF
Textdateien	- ASC, PDF/A
Datenbankdateien	- in ASC konvertieren
Videodateien	- keine Empfehlung
Audiodateien	- keine Empfehlung

Speichermedien

Disketten	- ungeeignet
ZIP-Disketten	- ungeeignet
Magnetbänder	- bedingt geeignet
Festplatten	- bedingt geeignet
CD	- bedingt geeignet
DVD	- bedingt geeignet
Blu-ray	- ungeeignet
USB Sticks	- ungeeignet

Keine Angst vor XML

Wer sich mit „Langzeiterhaltung digitaler Daten“ befasst, trifft über kurz oder lang auf das Kürzel „XML“. Die Abkürzung steht für „Extensible Markup Language“. Doch was ist XML? Die Bezeichnung „Language“ lässt vermuten, dass es sich um eine Computersprache handelt. Das W3C, das Internet-Konsortium, empfiehlt den Einsatz von XML und beginnt seine Erläuterungen mit: „Extensible Markup Language (XML) ist ein einfaches, sehr flexibles Textformat, abgeleitet von SGML (ISO 8879)“ (<http://www.w3c.org/XML/>). Demnach handelt es sich eher um ein Format als um eine Sprache.

XML ist dennoch kein Dateiformat. Ähnlich wie in HTML werden XML-Daten als einfacher Text gespeichert. Während einfache ASCII-Dateien den Text aneinanderreihen gibt XML die Möglichkeit, Text zu strukturieren und diese Struktur mit Auszeichnungen zu versehen. Beispielsweise kann ein Kennzeichen („Markup“) namens „<titel>“ in der XML-Textdatei enthalten sein. Die XML-Datei selbst enthält einen Verweis auf weitere Dateien, in welchen festgehalten ist, wie der Text zwischen <titel> und </titel> (letzteres kennzeichnet das Ende des Markup) dargestellt werden soll und wie der Text strukturiert ist. Solche zusätzlichen Dateien heißen Dokumenttyp-Deklaration (DTD) und Stylesheet-Definition.

Ein XML-Dokument beinhaltet auf diese Weise eine logische Struktur, beispielsweise:

[An erster Stelle steht, welche Dokumenttyp-Deklaration zu verwenden ist]

```
<titel>Nicht von Dauer</titel>
<untertitel>Kleiner Ratgeber</untertitel>
<gliederung>Inhaltsverzeichnis</gliederung>
<kapiteltitel>Liebe Kolleginnen, liebe Kollegen</kapiteltitel>
<text>eigentlich ...</text>
<kapiteltitel>Wie man es sieht</kapiteltitel>
```

Insbesondere für die Nutzung von XML-Dateien im Zusammenhang mit Datenbankdaten wurden Sets von Markups definiert, die international verwendet wird. Es gibt z.B. die so genannte Dublin-Core-Museum-Definition, kann beispielsweise folgende Form annehmen:

```
<?xml version="1.0" ?>
<dc-record>
<type>Physical object</type>
<type>original</type>
<type>cultural</type>
<format>L:81 W:52 H:2</format>
<title>Kite, insect, fly ? </title>
<description>Kite, L:81 W:52 H:2, collected in China by Berthold Laufer (1874-1934) in 1903,
```

```

depicting an insect, possibly a fly. Materials: paper, bamboo, pigment, string, masking tape
(modern), metal (modern). Native term: Ts'ang Yin [Shang-Yin?].</description>
<subject>Kites</subject>
<subject>Insect</subject>
<subject>Fly</subject>
<contributor>Laufer, Berthold</contributor>
<publisher>American Museum of Natural History Division of Anthropology</publisher>
<date>1903</date>
<identifier>AMNH-ANT 70/10596</identifier>
<relation>HasFormat CD269/CD269/70/10596.PCD</relation>
<relation>HasFormat http://anthro.amnh.org/images/full/70/10596.jpg</relation>
<relation>HasFormat http://anthro.amnh.org/images/preview/70/10596.jpg</relation>
<relation>HasFormat http://anthro.amnh.org/images/thumbnails/70/10596.jpg</relation>
<relation>HasFormat AMNH-LIB 7683</relation>
<relation>HasFormat AMNH-LIB 338999</relation>
<rights>American Museum of Natural History Division of Anthropology</rights>
</dc-record>

```

[entnommen aus: http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf]

Solcherart oder ähnlich strukturierte Objektbeschreibungen, welche standardisierte Markups verwenden, lassen sich leicht und vollautomatisch in Datenbanken einbinden - in eigene und andere.

Die vielfältigen Möglichkeiten von XML, seine einfache Struktur, die dennoch komplexe Darstellungen erlaubt, und seine wachsende Anerkennung als internationaler Standard lassen es als ein – auch (und besonders) aus der Sicht der „Langzeiterhaltung digitaler Daten“ - geeignetes Mittel zur Strukturierung von Textdateien erscheinen.

E-Mail-Archivierung

Im Hinblick auf die Archivierung von E-mails und andere digitale Medien gibt es unter internationalen Experten deutliche Bestrebungen, die Archivierung nicht als gesondertes intellektuelles Unterfangen anzusehen, sondern mehr als Teil eines größeren Vorgangs, der gemeinhin als „digitale Kuration“ bezeichnet wird. Aus diesem Grund beginnen Empfehlungen bezüglich der Aufbewahrung von E-Mails und anderen digitalen Medien bei den Datenerzeugern und nicht bei den Datenverwahrern. Archivare nehmen in dieser Konstellation die Rolle von Anwälten oder Beratern ein, die nicht nur Möglichkeiten definieren, wie digitale Medien archiviert werden, sondern auch, wie man sie erzeugen und gebrauchen sollte – heute und bis weit in die Zukunft. Die abgebenden Institutionen haben in diesem Prozess teil an der Arbeit des Archivs, da Aufgaben des Erzeugens, Verwahrens und der Aufbereitung von Dokumenten immer näher aneinander rücken.

Verfahrensweisen

Das vorrangige Ziel sollte es sein, den Gebrauch von E-Mails durch eine institutionsweite Verfahrensweise zu standardisieren, die grundlegende Archivierungs- und Übergabeverfahren bereits spezifiziert. Alle abgehenden E-Mails sind so zu gestalten, dass sie einem vorgegebenen Muster so weit wie möglich entsprechen. Aktuelle Best-Practice-Empfehlungen lassen sich wie folgt zusammenfassen: Vereinheitlichen Sie Überschriften und Unterschriften-Blocks von E-Mails, um ausreichend kontextuelle Metadaten zu erzeugen. Stellen Sie sicher, dass offizielle Nachrichten auf IMAP und nicht über POP3 Nachrichtenprotokolle gespeichert werden, damit wertvolle Dokumente nicht bei individuellen E-Mail-Nutzern oder auf persönlichen Heimcomputern stranden. Begrenzen Sie den Zugriff auf gespeicherte, d.h. bereits im Besitz des Repositoriums befindliche Nachrichten. Richten Sie innerhalb des Archivs einer Buchungskontrolle ein, um die Handlungen einer begrenzten und identifizierbaren Anzahl von Personen nachvollziehen zu können. Hierdurch stellen Sie sicher, dass keine Nachrichten mit Absicht im Nachhinein verändert werden.

Rechtliche Aspekte

Ein oft angeführter Grund, bezüglich der E-Mail-Verwahrung einem umfassenden Ansatz zu folgen, ist ein rechtlicher. Um die potentielle rechtliche Haftung eines Archivs zu ermessen, sollte man für jedes Repositorium eine Risikoabschätzung erarbeiten. Das Kernproblem ist hier die Trennung von privaten und dienstlichen E-Mails. In Deutschland sind Unternehmen und Verwaltungen verpflichtet, bei privaten Inhalten das Post- und Fernmeldegeheimnis zu wahren. In Deutschland finden des Weiteren die Bestimmungen

der „Abgabeordnung für die ordnungsgemäße Aufbewahrung von Unterlagen sowie für die Buchführung und weitere Aufzeichnungen“ Anwendung, die für digitale Dokumente in den sogenannten „Grundsätzen zum Datenzugriff und zur Prüfbarkeit digitaler Unterlagen“ (GdPdU) näher spezifiziert sind. Die GdPdU fordern explizit die Aufbewahrung in einem Format, das eine maschinelle Auswertbarkeit möglich macht. Nach dieser Anforderung können z.B. Excel-Tabellen nicht im eigentlich archivfähigen Format PDF/A aufbewahrt werden.

Obgleich tauglich für Attachments und kleine Textsammlungen, ist das PDF/A Format für die E-Mail-Archivierung ebenfalls nicht optimal geeignet, allein schon wegen des Volumens elektronischer Nachrichten, die im Laufe des nächsten Jahrzehnts in Archivsammlungen aufgenommen werden. Ebenso einschlägig ist das „Gesetz zur Regelung des Zugangs zu Informationen des Bundes“ (IFG). Andere rechtliche Regelungen zur Sicherung des geistigen Eigentums wie Firmenmarken, Patente, Copyright, Design und Rechte an Datenbanken (die sich in vielen Fällen bei jeder Änderung der Datenbank erneuern), können in Bezug auf einzelne Dokumente Anwendung finden.

Ein Überblick über diese und einschlägige EU-Richtlinien wird die Politik des Archivs bezüglich der Sperrfristen und der Trennung von persönlichen und inoffiziellen Nachrichten von offiziellen Dokumenten der Organisation leiten müssen. Werden E-Mails aus rechtlichen oder anderen Gründen vernichtet, so müssen auch alle Kopien dieser E-Mails auf allen Speicherträgern vernichtet werden, die aus Kontinuitätsgründen aufbewahrt werden.

Technische Ansätze

Nur ein kleiner Prozentsatz von E-Mails wird für die historische Forschung von Belang sein, und wie in rechtlicher Hinsicht ist auch hier das derzeit in Fachkreisen am meisten diskutierte Problem das der Authentizität. Informationsmanager sind sich einig darüber, dass die Authentizität von E-Mails am besten durch die Bewahrung von möglichst vielen Metadaten sicher gestellt wird. Sie betonen auch, dass Back-Up Träger bislang weder über eine Struktur verfügen, um den Kontext und die Beziehungen von Nachrichten nachzuvollziehen, noch um das Ausheben zu erleichtern. Es herrscht ein beträchtlicher Enthusiasmus für XML-basierte Ansätze vor, die auf das Open Archival Information System (OAIS) Referenz-Modell aufbauen. Allerdings gibt es derzeit noch keine Instrumentarien, mit denen das gesamte OAIS bedient werden kann. Ein großer Teil der Diskussionen dreht sich um die Erfordernisse von Authentizität, d.h. welche Attribute behalten werden müssen, um sicherzustellen, dass eine archivierte E-Mail auch auf ihre Authentizität hin geprüft werden kann. Die Kunst besteht darin, so früh wie möglich so viele Metadaten wie möglich festzuhalten – und das bedeutet im Moment der Entstehung.

So verführerisch es für Archive erscheinen mag, in ein schlüsselfertiges System zu investieren, ist eine solche Strategie jedoch derzeit keine optimale Lösung. Zum gegenwärtigen Zeitpunkt der Entwicklung von E-Mail-Archivierung einen kommerziellen Anbieter zu beauftragen, birgt erhebliches Risiko, ein System zu etablieren, das mittelfristig zur Informationsinsel wird. Wenn der Anbieter eines kommerziellen digitalen Speichersystems bankrott geht, kann auch das Archiv vor dem Aus stehen. In Bezug auf die Anforderungen an Lesbarkeit, Zugriffsmöglichkeit und Verwaltung kam 2006 eine belgische Archivstudie zu dem Ergebnis, dass hochproprietäre E-Mail-Archivierungssysteme gegenüber regulären E-Mail Programmen (wie etwa Lotus Notes) keinen Mehrwert erbringen. Im Lichte dieser Schwierigkeiten und im Hinblick auf das vorrangige Ziel von Archiven, den Zugriff auf digitale Quellen über viele Jahrzehnte in die Zukunft hinein sicherzustellen, sollten sie offene Software implementieren, die den Export von E-Mails aus proprietären E-Mail-Systemen in eine möglichst nicht-proprietäre dauerhafte Archivierungslösung ermöglicht.

Ist die institutionsweite Politik des E-Mail-Managements etabliert, würde eine Zwischenlösung sich zunächst darauf konzentrieren, die originalen E-Mail-Übermittlungsdateien RFC 2822 zu speichern. Diese Dateien enthalten die Metadaten, die nötig sind, um zu belegen, dass die Nachricht vom angegebenen Absender zum angegebenen Rezipienten geschickt worden ist und halten fest, an welchem Tag und zu welcher Zeit die Nachricht bei den verschiedenen Servern aus- und eingegangen ist. Die Log-Dateien des E-Mail-Verkehrs sollten ebenfalls in diesem Interim-Deposit verwahrt werden, um zu dokumentieren, wann welche Nachricht an wen geschickt oder von wem empfangen worden ist. RFC 2822 kodiert Attachments.

Bei diesen Attachments handelt es sich um die unterschiedlichsten Formate, die für die Langzeitarchivierung jeweils spezifisch behandelt werden müssen. Entscheidend ist hier, die Attachments zu dekodieren, bevor sie in den Archivspeicher eingehen, und diese Dateien soweit wie möglich in nicht-proprietäre oder sogenannte offene (d.h. Definition freigegebener) Formate zu transferieren. Wenn dies geschehen ist, muss jedes Attachment wieder mit den entsprechenden RFC 2822-Dateien verknüpft werden. (Die Tatsache, dass man Links zwischen RFC-Dateien und einer beeindruckend langen Liste von angehängten Medien aufrecht erhalten muss, spricht ebenfalls dagegen, ein Document Management System zu kaufen und zu unterhalten, das ausschließlich E-Mails archiviert.) RFC 2822 Metadaten stellen auch eine gute Basis zur Handhabung der rechtlichen Risikoabschätzung dar (vgl. vorherigen Abschnitt).

Sind diese Zwischenschritte abgeschlossen, sollte man die besten Open Source XML-Lösungen testen. Zur Zeit stellt die XML-Variante des australischen Nationalarchivs eine solide Lösung mit der längsten Testlaufzeit dar. Eine viel versprechende neuere Ergänzung ist der vom Rockefeller Archive Center und dem Archiv der Smithsonian

Institution gemeinsam entwickelte sogenannte "Parser": Die Aufgabe dieses Computerprogramms ist es, Gruppen von E-Mails in ein XML-Format zu migrieren, das E-Mails in situ erfasst, d.h. vollständig mit Attachments und in dem organisatorischen Kontext, in dem sie vom Inhaber des E-Mail Accounts verwaltet wurden.

Wenn man potentielle Systeme für die Langzeitarchivierung von E-Mails testet, sollten Überlegungen die Diskussion von Zugriffsrechten und zukünftigen Darstellungsmodi der E-Mails bestimmen, wie die Inhalte von zukünftigen Forschern genutzt werden können. Die Bedürfnisse dieser Nutzer sollten auch die Wahl der deskriptiven Metadaten bestimmen, dem Counterpart der administrativen und konservatorischen Metadaten, die in den RFC 2822-Dateien enthalten sind. Dieses Stadium der Planung wird auch die rechtliche Evaluation des Zugriffs der Sammlung mittels standardisierter Nutzungsbedingungen erfordern. Im Ergebnis wird das Archiv den Nutzern eine Einverständniserklärung zur Datennutzung per Mausklick abverlangen und einen Link mit dem Wortlaut der entsprechenden rechtlichen Vorgaben des Archivs bereit stellen.

Ausgewählte Veröffentlichungen:

Filip Boudrez (2006), "Filing and Archiving E-Mail," retrieved 17 October 2009 from http://www.expertisecentrumdavid.be/docs/emailrapport_lr.pdf

Maureen Pennock, (July 2006), "Curating E-Mails: A Life-Cycle Approach to the Management and Preservation of E-mail Messages," DCC Digital Curation Manual, S. Ross, M. Day (eds.)
<http://www.dcc.ac.uk/resource/curation-manual/chapters/curating-emails>

Karin Schwarz (June 2009), "E-Mail-Archivierung," nestor-Handbuch, herausgegeben von Heike Neuroth, et. al.
<http://nestor.sub.uni-goettingen.de/handbuch/>

"Developing a Policy for Managing E-Mail," (2004), The National Archives (ed.)
http://www.nationalarchives.gov.uk/documents/managing_emails.pdf

Pilotprojekte:

Dutch National Archives e-mail preservation testbed, Xmail:
<http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=273>

National Archives of Australia's XENA e-mail project:
<http://xena.sourceforge.net/>

Rockefeller Archive Center and Smithsonian Institution Archives' e-mail preservation
parser:
<http://siarchives.si.edu/cerp/parserdownload.htm>

nestor und mehr

Weitere Informationen, einschließlich Expertisen, Erläuterungen zu Rechtsfragen und Informationen zu geplanten Veranstaltungen und Workshops, können im Internet abgerufen werden, bei:

„nestor“

Die Kurzbezeichnung „nestor“ steht für „Network of Expertise in long-term STorage of digital Resources“. Dahinter verbirgt sich ein Kooperationsverbund mehrerer Institutionen aus dem gesamten Kulturbereich in Deutschland. Die Partner sind:

Die Deutsche Bibliothek, Frankfurt und Leipzig
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Bayerische Staatsbibliothek, München
Humboldt-Universität, Berlin
Institut für Museumsforschung, SMB-PK, Berlin
Bibliotheks-Service Zentrum, Baden-Württemberg
Landesarchiv Baden-Württemberg
Fernuniversität Hagen
Institut für deutsche Sprache

- <http://www.langzeitarchivierung.de>

PADI

National Library of Australia, Preserving Access to Digital Information (PADI):

- <http://nla.gov.au/padi>

NDIIP

National Digital Information Infrastructure Program, US Library of Congress:

- <http://www.digitalpreservation.gov>

DPC

Die Digital Preservation Coalition (DPC) wurde 2001 ins Leben gerufen um die Anstrengungen zur Langzeitarchivierung zu bündeln.

- <http://www.dpconline.org>

Fragen und Anmerkungen bitte an:

Dr. Stefan Rohde-Enslin
mail: s.rohde-enslin@smb.spk-berlin.de

oder

Dr. Keith R. Allen
mail: kra@keithrallen.com