

1 Langzeitarchivierung digitaler Dokumente

Dieses Kapitel führt in die Problematik der Langzeitarchivierung digitaler Dokumente ein. Abschnitt 1.1 diskutiert Probleme der Haltbarkeit elektronischer Speichermedien sowie die Abhängigkeit digitaler Dokumente von spezifischer Software zu ihrer Interpretation. Abschnitt 1.2 dringt weiter zu den Ursachen dieser Probleme vor und erläutert relevante Begriffe und Konzepte. Grundlage aller Erhaltungsmaßnahmen für digitale Dokumente ist die in Abschnitt 1.3 beschriebene Bewahrung der zugrunde liegenden Zeichenfolgen und Metadaten. Abschnitt 1.4 erläutert die verschiedenen technischen Ansätze zur Archivierung im Zusammenhang. Die wichtigsten dieser Ansätze werden in den folgenden Kapiteln vertieft dargestellt. Juristische und gesellschaftliche Fragen, die auf dem Weg zu einer tragfähigen Lösung der Archivierungsproblematik digitaler Dokumente beantwortet werden müssen, sind Gegenstand von Abschnitt 1.5.

1.1 Vorzüge und Nachteile digitaler Medien

Digital documents last forever – or five years, whichever comes first. (Rothenberg 1999)¹

Bibliotheken, Archive und ähnliche Institutionen dienen der Erhaltung von Dokumenten mit bleibendem wissenschaftlichem, künstlerischem oder gesellschaftlichem Wert, sodass künftige Generationen ebenso wie wir heute darauf zugreifen können. Als erhaltenswert betrachten wir neben dem gesamten aktuellen Buchangebot auch wertvolle Originalmanuskripte, Ton- und Bildträger sowie in zunehmendem Umfang digitale Datenträger.

Während alte Folianten teure Konservierungsmaßnahmen erfordern und auf säurehaltiges Papier gedruckte Werke durch Spei-

Dokumente erhalten

*Vorzüge digitaler
Dokumente*

¹ siehe <http://www.clir.org/pubs/reports/rothenberg/introduction.html>.

cherung auf Mikrofilm vor dem endgültigen Untergang bewahrt werden müssen, scheinen sich digitale Medien durch ideale Eigenschaften auszuzeichnen: Bitfolgen lassen sich im Prinzip über längere Zeiträume ohne Informationsverlust aufbewahren. Die rasant steigende Speicherkapazität² erlaubt es schon heute, umfangreiche Nachschlagewerke auf einer einzigen CD zu speichern. Bei voranschreitender Miniaturisierung der Datenträger scheint sich mit der digitalen Speicherung eine attraktive, universell einsetzbare Basis der Archivierung anzubieten, die noch dazu weitere Vorteile mit sich bringt: Nicht nur die in den elektronischen Katalogen gespeicherten Angaben, sondern auch den gesamten Inhalt digitaler Dokumente können wir in Sekundenschnelle durchsuchen lassen. Ebenso rasch bringt das Internet Dokumente aus aller Welt an unseren Arbeitsplatz oder nach Hause, wo wir sie mit vielfältigen Werkzeugen bearbeiten und zu neuen Dokumenten zusammensetzen können.

Wo liegt dann das Problem?

nur indirekt lesbar

Wer über einen längeren Zeitraum hinweg Computer zur Textverarbeitung genutzt hat, kennt die (mehrteilige) Antwort: Zunächst einmal sind elektronische Datenträger für Menschen nicht unmittelbar lesbar. Während wir Jahrtausende alte Höhlenmalereien mit etwa dem gleichen Sinnesapparat wahrnehmen können wie die Zeitgenossen des Künstlers, erschließt sich uns der Inhalt der von uns selbst erzeugten Textdateien nur indirekt mit Hilfe eines geeigneten Computersystems. Für sich genommen ist das kein Problem. Auch der Tatsache, dass Speichermedien wie Disketten, Festplatten und CDs altern und ausfallen können, lässt sich mit geeigneten Maßnahmen (Sicherheitskopien) so begegnen, dass kein Datenverlust eintritt.

*Alterung durch
technischen Wandel*

Schneller als die Datenträger altern, veralten aber die Komponenten der Computersysteme: Zu 5,25-Zoll- oder gar 8-Zoll-Disketten finden sich heute kaum mehr geeignete Abspielgeräte und selbst wenn, dann ließen sie sich mangels neuer Treibersoftware nicht an moderne Computer anschließen. Noch größer sind Vielfalt und Wandel bei der Software. Im Laufe der Jahre entstanden tausende verschiedener Editoren für digitale Dokumente (Text-, Graphik-, Tabellenkalkulations-, Folien-, Webseiten-Editoren etc.), die meisten von ihnen mit jeweils eigenen internen Datenformaten. Wegen der Unverträglichkeit dieser Formate ist es häufig gar

²Infolge des »Gesetzes von Moore« verdoppeln sich die Verarbeitungsgeschwindigkeit und der Speicherumfang von Computern etwa alle 18 Monate.

nicht oder nur unter erheblichem Qualitätsverlust möglich, Dokumente zwischen solchen Editoren auszutauschen.

Langfristig überstehen nur relativ wenige dieser Systeme den Wettbewerb im Markt. Nutzer kommerziell unterlegener Produkte stehen am Ende vor dem Problem, wie sie ihren Bestand veralteter Dokumente in eine neue Umgebung retten können. Mit arbeitsaufwändigen, »heroischen Maßnahmen« wie der Neuerstellung der wichtigsten Dokumente oder dem Einscannen von Texten in neue Systeme (mit anschließender manueller Nachbearbeitung) gelingt dies in der Regel. Offensichtlich ist diese Vorgehensweise nicht für die riesigen Dokumentenbestände von Bibliotheken und Archiven geeignet, in denen mit wenig Personal auf Dauer der Fortbestand und der Dokumentenzugriff gesichert werden müssen.

aufwändige Erhaltungsmaßnahmen

1.2 Herausforderungen, Begriffe und Konzepte

Computer speichern Daten als Folgen von Nullen und Einsen, d. h. als so genannte *Bitströme*. Auch digitale Dokumente sind im Kern nichts anderes als solche Bitströme. Wir sprechen etwas allgemeiner von *digitalen Zeichenströmen*, da man statt mit Nullen und Einsen ebensogut mit anderen Basiszeichen arbeiten kann. Auf die Eigenschaften digitaler Zeichenströme und ihre Konservierung werden wir in Abschnitt 1.3 eingehen.

digitale Dokumente sind Bitströme

Für den Menschen ist der reine Datenstrom eines digitalen Dokuments ohne ein passendes *Abspielsystem* nicht zugänglich. Erst in einer geeigneten Umgebung werden aus den Zeichenfolgen wieder die Texte, Graphiken, Bewegungen, Töne und weiteren Bausteine, aus denen sich ein multimediales Dokument zusammensetzt. Häufig ist die einzig geeignete Umgebung das Editorprogramm, mit dem das Dokument ursprünglich erstellt wurde bzw. eine hinreichend ähnliche Version dieses Programms. Wie schon erwähnt, unterscheiden sich die Editoren verschiedener Softwarehersteller und in geringerem Maße auch schon die verschiedenen Versionen des gleichen Editorprogramms in der Art und Weise, wie sie digitale Dokumente strukturieren und als Datenströme speichern.

Zugang über Abspielsysteme

Betrachten wir als typisches Beispiel zwei einfache, hypothetische Texteditoren, T1 und T2: T1 speichert einen Text als Folge von Zeilen und jede Zeile als Folge von Zeichen. Zu jedem Zeichen legt T1 weitere Informationen ab, u. a. den Zeichensatz, die

Beispiel: unterschiedliche Textformate

Größe des Zeichens, ob es fett und/oder kursiv ist und wie viele Bildpunkte oberhalb bzw. unterhalb der Zeilengrundlinie sich das Zeichen befindet. Ganz anders die Speicherung von Texten bei T2: Hier wird der gesamte Text als eine Folge von Textzeichen gespeichert, in die so genannte *Steuerkommandos* eingestreut sind. Beispiele für solche Steuerkommandos sind Absatzende, fett ein, kursiv aus, verwendeZeichensatz(font), hochstellen, tiefstellen, auf Zeile und verwendeFarbe(color). Während die Entwickler von T1 eine möglichst flexible Positionierung der Zeichen auf der Textseite anstrebten, ließen sich die Entwickler von T2 eher von den Eigenschaften eines Druckers leiten. T1-Dokumente sind schwarzweiß und können wegen der punktgenauen Zeichenpositionierung auch komplizierte wissenschaftliche Formeln enthalten. T2-Dokumente dagegen sind weniger für Formeln geeignet, dafür aber bunt und kompakter in der Speicherung.

Konversion zwischen
diesen Formaten

Die oben skizzierten Interna von T1 und T2 bezeichnet man als ihre *Datenformate*. Wegen der unterschiedlichen Datenformate kann man T1-Dokumente offenbar ebenso wenig mit T2 betrachten wie T2-Dokumente mit T1. Aus kommerziellen Gründen (Abwerbung von Kunden) bieten Hersteller von Texteditoren häufig Konversionsprogramme an, mit denen sich Dokumente eines fremden Datenformats in das eigene Datenformat umwandeln lassen. Man spricht von so genannten *Importfiltern*. Beispielsweise würden die Hersteller von T1 einen Importfilter für T2-Dokumente anbieten, um T2-Nutzern den Übergang zum eigenen System zu erleichtern. Dennoch müssen T2-Nutzer Informationsverlust hinnehmen, wenn sie den Umstieg auf T1 vornehmen: Da T1-Dokumente stets schwarzweiß sind, gehen die in T2-Dokumenten eventuell vorhandenen Farbunterscheidungen verloren. (Auch die umgekehrte Konversion von T1 nach T2 wäre zwangsläufig mit Informationsverlust verbunden, da sich die feine Zeichenpositionierung von T1 nur grob mit T2 annähern ließe.)

Filter und
standardisierte
Austauschformate

Wegen der unbegrenzten Vielfalt der Datenformate werden Importfilter jeweils nur für einige wenige, gängige Dokumentenformate erstellt. Anbieter weniger populärer Editoren stellen manchmal auch *Exportfilter* bereit, mit denen sich Dokumente aus dem eigenen in ein gängigeres Format umwandeln lassen. Damit geben sie den eigenen Kunden die Sicherheit, die erstellten Dokumente auch dann weiter verwenden zu können, wenn der eigene Editor sich wirtschaftlich nicht behaupten sollte. Angesichts dieses Wildwuchses an Datenformaten liegt es nahe, an *Standardisierung* zu denken oder zumindest ein gemeinsames Austauschformat zu schaffen: Theoretisch hätte man die vollständige Kon-

vertierbarkeit zwischen beliebigen Datenformaten erreicht, wenn es ein solches *Austauschformat* AF und zu jedem Editor je einen Import- und einen Exportfilter zwischen AF und dem eigenen Format gäbe. Man macht sich leicht klar, dass dann aber alle Editoren (anders als oben T1 und T2) die gleichen oder wenigstens äquivalente Ausdrucksmöglichkeiten besitzen müssten (vgl. Abb. 1.1).

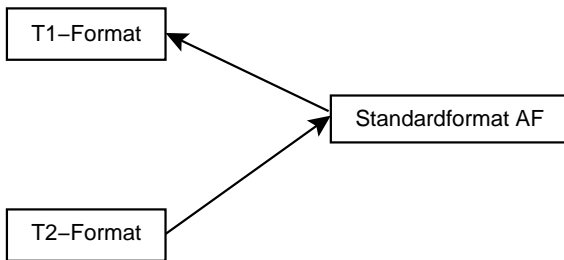


Abbildung 1.1
Austauschbarkeit
über Standardformate

Das Haupthindernis bei einem Standardisierungsansatz ist paradoxerweise der Fortschritt, da er immer neue Möglichkeiten mit sich bringt. Zu reinen Textformaten sind in relativ kurzer Zeit viele neue Formen hinzugekommen: Graphik, Sound, Video, Hypertext, Spreadsheets, jede einzelne dieser Formen wiederum in vielen proprietären Datenformaten. Zusätzlich werden diese Formen in allen denkbaren Kombinationen zu elektronischen Multimediadokumenten verbunden. Eine Konsolidierung dieser Entwicklung oder zumindest eine Verlangsamung der Innovationsrate ist nicht in Sicht. Solchen anhaltenden Veränderungen kann ein einmal gewähltes, festes Datenformat auf Dauer grundsätzlich nicht gerecht werden. Einen Ausweg aus diesem Dilemma bieten Sprachen, in denen Datenformate beschrieben und erweitert werden können. Dazu gehören die in Kapitel 5 behandelten *Auszeichnungssprachen* wie SGML und XML. Auf Basis dieser Standards werden derzeit ergänzende Sprachen und Werkzeuge entwickelt, mit denen man die systematische Transformation von Dokumenten aus einem Format in andere beschreiben und durchführen kann.

*Fortschritt als
Hindernis*

ein möglicher Ausweg

Die neuen Datenformate unterscheiden sich von ihren Vorgängern nicht nur durch ihre Vielfalt, sondern auch durch eine reichere konzeptionelle Strukturierung: In alten Systemen (wie T1 und T2) hat man beispielsweise Tabellen, mathematische Formeln und sogar Liniengraphiken wie gewöhnlichen Text aus Buchstaben und Symbolen zusammengesetzt und für das menschliche Auge durch Einstreuen von Zwischenräumen notdürftig erkennbar gemacht. Im Unterschied zu neueren Textformaten, die *Konzepte* wie

*Wieso immer neue
Formate?*

Tabellen und Formeln als Bausteine bereithalten, wurden früher die gleichen Konzepte nicht explizit dargestellt, sondern wie beschrieben durch eine Ad-hoc-Formatierung nachgebildet. Die Lesbarkeit solcher Nachbildungen hängt übrigens wesentlich von einer scheinbar nebensächlichen Eigenschaft früher Editoren ab: Es gab beim Drucken nämlich nur Zeichensätze fester Breite. Abbildung 1.2 zeigt eine einfache Liniengraphik und ihre Darstellung durch das für dieses Buch verwendete Textsystem \LaTeX (in Fließtext fasst \LaTeX Zwischenraumfolgen und einzelne Zeilenwechsel zu je einem Zwischenraum zusammen):

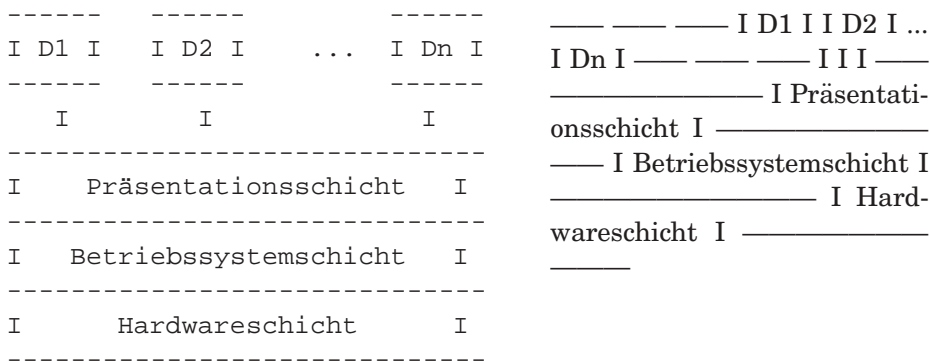


Abbildung 1.2 Eine Liniengraphik und was \LaTeX daraus macht

*Probleme bei der
Konversion von
Texten*

*neue Formate neue
Probleme*

*Was ist ein
»Abspielsystem«?*

Hardware

Da ein Stillstand bei der Einführung neuer Konzepte nicht zu erwarten ist, ergibt sich also ein weiteres Problem: Die Konversion von Texten in neuere, konzeptionell reichere Formate erfordert das Erkennen von Konzepten, die durch Ad-hoc-Formatierung nur implizit gegeben sind. Das setzt ein gewisses inhaltliches Verständnis des Textes voraus und ist daher nur schwer, wenn überhaupt durch ein Programm automatisch und verlässlich zu erledigen.

Bislang haben wir »Abspielsystem« gleichgesetzt mit dem Editor- oder Anzeigeprogramm, mit dem das Dokument am Bildschirm betrachtet werden kann. Das ist aber nur der Kern einer funktionsfähigen Abspielumgebung, die daneben mindestens auch die Hardware und Betriebssoftware eines passenden Computers, eventuell auch zusätzliche Geräte und zugehörige Treiberprogramme umfasst. Die obige Liniengraphik stellt den Gesamtaufbau einer Abspielumgebung schematisch dar:

- Zur Hardwareeschiicht gehört neben dem »nackten« Computer mit Prozessor, Arbeitsspeicher und Busverbindungen ei-

ne Reihe von notwendigen oder wünschenswerten Ergänzungen wie Graphikkarte und Bildschirm, sekundäre, dauerhafte Speicher (Festplatte, Magnetband, optische Medien etc.) und Eingabegeräte (u. a. Tastatur, Maus, Joystick, Mikrofon).

- Das Betriebssystem und die Treiberprogramme bilden zusammen eine Schicht von Programmen, die es dem menschlichen Nutzer erlauben, von den für ihn weniger relevanten, physikalischen Eigenschaften der Hardware zu abstrahieren und sich auf die so genannten »logischen« Eigenschaften des Rechensystems zu konzentrieren. Beispielsweise sieht der Nutzer die Festplatte nicht als eine in Ebenen, Segmente und Spuren gruppierte Menge von Magnetisierungspunkten, sondern als ein Dateisystem mit ineinander geschachtelten Verzeichnissen, welche Dateien mit elektronischen Dokumenten enthalten.
- Die Programme der Präsentationsschicht (u. a. Texteditor, Zeichenprogramm, Web-Browser) setzen auf den darunter liegenden Schichten auf und sorgen für eine geeignete Anzeige der Dokumente D1, D2, . . . , Dn. Welches der Präsentationsprogramme geeignet ist, hängt vom Typ des elektronischen Dokuments ab (z. B. Word2000-Datei, TIFF-Graphik, HTML-Datei).

Systemsoftware

Präsentationssoftware

Ein Abspielsystem ist also kein monolithischer Block, sondern wird aus einer Vielzahl von Komponenten zusammengesetzt, die in komplizierter Weise miteinander interagieren.

Ebenso wie die Präsentationsprogramme sind die anderen Komponenten dieser Systeme einem steten Wandel ausgesetzt: Laufend werden neue, bessere Gerätetypen und -varianten entwickelt. Damit diese genutzt werden können, bedarf es neuer Systemsoftware-Komponenten, eventuell sogar Anpassungen der Präsentationsprogramme. Trotz einer gewissen Marktberreinigung bei den Betriebssystemen (hin zu Windows und Unix) gibt es auch davon eine Unmenge von Varianten und Versionen. Aufgrund anhaltender Innovationsschübe – besonders eindrucksvoll darunter die Einführung graphischer Benutzungsoberflächen und das Entstehen des Internets – wird Software auch ohne physikalische Abnutzung regelmäßig nach wenigen Jahren obsolet und ist durch verbesserte Versionen oder neuere Produkte zu ersetzen. Da neue Versionen von Anwendungsprogrammen natürlich nur für aktuelle Betriebssystemversionen entwickelt werden, bedingen sich Erneuerungen von Betriebssystem und Anwendungssoftware häufig

*Aktualisierungsdruck
bei Abspielsystemen*

gegenseitig. Die Einbindung von immer mehr Rechnern ins Internet verstärkt diesen Aktualisierungsdruck.

Folgeprobleme

Die oben am Beispiel von Texteditoren beschriebenen Herausforderungen bei der Archivierung elektronischer Dokumente potenzieren sich insgesamt, weil

- ❑ wie dargestellt Änderungen an einer Komponente der Abspielumgebung Änderungen an den anderen Komponenten nach sich ziehen;
- ❑ bei Multimediadokumenten verschiedene Präsentationsprogramme in einer (Browser-)Umgebung interagieren und daher zusammenpassen müssen.

bibliographische Metadaten

Bei der elektronischen Archivierung sind neben dem eigentlichen Dokument eine Reihe von Angaben über das Dokument, so genannte *Metadaten*, aufzubewahren. Dabei handelt es sich zum Teil um bibliographische Angaben, wie man sie schon in herkömmlichen Bibliothekskatalogen findet, z. B.:

- ❑ Autor, Titel, Verlag, Erscheinungsjahr und evtl. die ISBN-Nummer: Angaben, die das Dokument eindeutig identifizieren.
- ❑ Fachgebiet und Schlagwörter: Angaben, die neben Autor und Titel für die Dokumentenrecherche benötigt werden.
- ❑ Speicherort (analog zu Aufstellungsort): Angabe, wo man das Dokument tatsächlich findet.

Metadaten für die Langzeitarchivierung

Weitere Metadaten sind spezifisch für die elektronische Langzeitarchivierung:

- ❑ Codierung und Datenformat: Angaben, die zum Lesen des Dokuments unbedingt erforderlich sind.
- ❑ »Migrationshistorie« bzw. Beschreibung der Originalumgebung: Angaben, welche auf den Originalzustand des Dokuments schließen lassen.
- ❑ Copyright-Vermerke u. ä.: Angaben, welche die legalen und sonstigen Konditionen des Zugriffs auf das Dokument beschreiben.

Der Hintergrund zu den letzten drei Angaben wird in den Abschnitten 1.3–1.5 erläutert.

Metadaten beim Dokument speichern?

Elektronische Dokumente und ihre Metadaten gehören einerseits eng zusammen, da das eine ohne das andere wertlos ist und daher beide unbedingt konsistent gehalten werden müssen. Andererseits gibt es Gründe, beide getrennt voneinander zu speichern:

Metadaten werden u. a. verwendet, um Dokumente aufzufinden. Das geht schneller, wenn Kataloginformationen gesondert aufbewahrt werden. Noch gewichtiger ist das Argument, dass der Dokumenteninhalte überhaupt erst dann gelesen werden kann, wenn man anhand der Metadaten die Codierung und das Datenformat des Dokuments bestimmt hat.

Die Frage der Speicherung und Kapselung von elektronischen Dokumenten und Metadaten ist grundlegend für die Organisation und den Betrieb eines elektronischen Archivs, ebenso die Struktur der von dem Archiv laufend durchzuführenden Prozesse (wie Aufnahme eines Dokuments, Speicherung und Auffrischung, Ausleihe von Dokumenten). Mit dem vom *Consultative Committee for Space Data Systems* (CCSDS) im Auftrag der ISO entwickelten OAIS-Referenzmodell³ erhält man hierzu eine allgemein akzeptierte, in den Details aber noch auszugestaltende Vorlage.⁴ Das OAIS-Referenzmodell ist Gegenstand von Kapitel 2.

elektronische Archive

1.3 Konservierung von Zeichenströmen

Grundlage einer jeden Archivierung digitaler Dokumente ist die Konservierung von Zeichenströmen. Bevor wir in Abschnitt 1.4 auf die unterschiedlichen technischen Ansätze zur Langzeitarchivierung eingehen, soll hier die gemeinsame Grundlage aller Verfahren dargestellt werden. Ausgangspunkt ist die folgende Erklärung:

Erklärung: Ein *digitales Dokument* besteht aus dem digitalen Dokumenteninhalte und den zum Dokument gehörigen Metadaten. Der *digitale Dokumenteninhalte* ist als ein Strom (Folge) von Zeichen eines endlichen Zeichensatzes A codiert. Zu den Metadaten gehört die Angabe des zur Codierung des Dokumenteninhalts verwendeten Zeichensatzes A sowie weitere Angaben (u. a. Autor, Titel). Die Metadaten sind als Folgen eines (Basis-)Zeichensatzes B codiert. Die Zeichensätze A und B können, müssen aber nicht übereinstimmen. Die Struktur der Metadaten und der zugrunde liegende Zeichensatz B sind archivabhängig festgelegt.

Was ist ein »digitales Dokument«?

Das Adjektiv »digital« weist auf die wesentliche Eigenschaft hin, nämlich die Codierung von Inhalten als Zeichenfolgen. Betrachten wir als Beispiel ein Graphikdokument, das u. a. eine gerade

z. B. PostScript

³Das Akronym OAIS steht für *Open Archival Information System*.

⁴siehe <http://www.ccsds.org/RP9905/RP9905.html>.

Linie vom Punkt mit den Koordinaten (12.4, 3.1415) zum Punkt (202.4, 47.11) enthält. Liegt das Graphikdokument im PostScript-Format vor, dann enthält der digitale Dokumenteninhalte u. a. die Zeichenfolge

```
12.4 3.1415 moveto 202.4 47.11 lineto
```

die textuell ausdrückt, dass sich der Zeichenstift in gehobenem Zustand zum Punkt (12.4, 3.1415) bewegen und von dort auf dem Papier zum Punkt (202.4, 47.11) fahren soll. (Ein »PostScript-fähiger« Laserdrucker ist in der Lage, diese textuelle Beschreibung in eine gerade Linie umzusetzen.)

Standardcodes wie
ASCII und Unicode

In der Praxis häufig verwendete Zeichensätze sind die von Standardcodes wie ASCII, EBCDIC, Unicode und UCS. Solche Codes enthalten eine feste, endliche Menge von Zeichen und legen eine Codierung dieser Zeichen durch Bitfolgen fest. Die Codes ASCII und EBCDIC verwenden dabei Bitfolgen der Länge 8, Unicode Bitfolgen der Länge 16 und UCS sogar Bitfolgen der Länge 32. Während ASCII an Zeichen nur die Groß- und Kleinbuchstaben, die arabischen Ziffern, die Interpunktionszeichen, mathematische Grundsymbole wie + und < und etwa noch mal so viele Sonderzeichen enthält, deckt Unicode die Schriftzeichen der wichtigsten lebenden Sprachen ab. Mit UCS wird es künftig möglich sein, alle bekannten Schriftzeichen zu codieren (vgl. hierzu Rechenberg und Pomberger (1997), Kapitel B1). Beispiel: Die ASCII-Codierung des Zeichens »m« ist die Bitfolge 01101101, die des Zeichens »5« ist 00110101.

»binäre«
Datenformate

Erklärung: Ein *Datenformat* heißt *binär*, wenn die zugehörigen Dokumenteninhalte nur Folgen von Nullen und Einsen sein können. Wenn der Dokumenteninhalte aus Zeichen eines anderen Zeichensatzes (z. B. dem des ASCII-Codes) besteht, dann spricht man von einem *textbasierten* Datenformat.

In diesem Sinne ist das oben verwendete PostScript-Format ein textbasiertes Datenformat. Bei elektronischer Speicherung eines solchen Textes wird entsprechend dem Code die Zeichenfolge in eine Bitfolge umgewandelt und so letztlich auch binär gespeichert.

Wo werden digitale
Dokumente
gespeichert?

Derzeit werden elektronische Dokumente häufig als *Dateien* eines Rechensystems abgelegt. Andere Möglichkeiten sind die Speicherung in einer *Datenbank* oder in einem dedizierten *Dokumentenverwaltungssystem*. Schließlich gibt es so genannte *dynamische Dokumente*, deren Datenströme erst beim Zugriff auf das Dokument entstehen, indem aktuelle Daten bei Bedarf nach

einem vorgegebenen Schema (per Programm) zusammengestellt werden.

Ebenso wie Tontafeln, Pergament, Papier und andere physikalische Träger von Dokumenten sind elektronisch lesbare Datenträger wie Magnetplatten und -bänder und auch optische Speichermedien wie CDs nur von begrenzter Haltbarkeit. Selbst bei regelmäßiger Auffrischung des Inhalts ist der ursprüngliche Datenträger nach wenigen Jahren unbrauchbar. Wie schon dargestellt, liegt ein noch größeres Problem darin, dass der technische Fortschritt laufend bessere Speichermedien hervorbringt und Abspielgeräte für den jeweils aktuellen Standard nur relativ kurze Zeit verfügbar sind. Die in der Datenverarbeitung allgemein bekannte Konsequenz ist, dass sich elektronische Daten und Dokumente nur durch fortgesetztes Kopieren und durch Migration von Speichermedium zu Speichermedium erhalten lassen. Kapitel 3 stellt Migrationsverfahren für elektronische Dokumente ausführlich dar.

*fortgesetztes
Kopieren als
Erhaltungsstrategie*

Die für die Codierung von Texten verwendeten Codes (ASCII etc.) haben sich als verhältnismäßig dauerhaft erwiesen. Außerdem kann man auf älteren Codes basierende Datenströme so in Unicode- oder UCS-basierte Datenströme umwandeln, dass die ursprünglichen Datenströme bei Bedarf bit-identisch wiederhergestellt werden können. *Aus heutiger Sicht erlauben Maßnahmen wie fortdauerndes Kopieren, Migration und reversibler Übergang zu neuen Codes eine sichere elektronische Konservierung von Zeichenströmen über lange Zeiträume hinweg.* Eine zusätzliche Speicherung auf haltbaren Medien (Papier, Mikrofilm, Metall o. ä.) kann die Sicherheit des Erhalts von Zeichenströmen weiter erhöhen. Neben dem Übergang zu neueren Codes werden elektronische Datenströme gelegentlich weiteren Transformationen unterworfen, z. B. der Datenkompression zum Zweck kompakter Speicherung oder der Verschlüsselung vertraulicher Dokumenteninhalte. Sofern solche Transformationen eindeutig (und per Programm) umkehrbar sowie in den Metadaten ausreichend dokumentiert sind, beeinträchtigen sie die Konservierung des Zeichenstroms nicht.

*Transformationen wie
Datenkompression
und Verschlüsselung*

1.4 Technische Ansätze zur Langzeitarchivierung

Aufgabe: Formate interpretieren

Während sich die Konservierung von Zeichenströmen als lösba-re technische Aufgabe herausstellt, ist damit für das Problem der elektronischen Langzeitarchivierung nur eine wichtige Voraussetzung geschaffen. Es bleibt die Aufgabe, die ungeheuer vielfältigen, auf binären oder textbasierten Zeichenströmen aufsetzenden Datenformate von Multimediadokumenten zu interpretieren, d. h., auch nach langer Zeit die textuellen, graphischen und akustischen Bestandteile solcher Dokumente wieder in der »gleichen Form« darzustellen, die schon ihre Erzeuger wahrgenommen haben. Bei interaktiven Dokumenten wie Hypertext, Animationen oder Spreadsheets kommt die Forderung nach »gleichartiger Reaktion« auf Benutzereingaben hinzu.

Wie war das bei Tonträgern?

Einige *Analogien zur Langzeitarchivierung elektronischer Dokumente* bietet das Feld der Schallkonservierung, das mit Edisons Phonographen 1877 seinen Anfang nahm. Das Prinzip, Schallwellen in wellenförmige Auslenkungen einer Rille auf dem Tonträger umzusetzen, hat sich – unter erheblichen Modifikationen – bis zu heutigen Schallplatten erhalten. Die Verbesserung von Tonqualität, Haltbarkeit und Aufnahmedichte motivierten die Wechsel des Tonträgermaterials von Stanniol über eine Wachsmasse zu Schellack-Harz und Vinyl-Kunststoff. Weitere Technologiesprünge sind verbunden mit elektrischer Wiedergabetechnik, Stereo- und Quadrophonie sowie HiFi-Technik. Daneben entwickelten sich Alternativen, vornehmlich die elektromagnetische Speicherung auf Magnetbändern, die es dem Benutzer erlauben, eigene Aufnahmen zu erstellen und vorhandenes »Tonmaterial« zu kopieren. Mit der Einführung von Digitalaufnahmen hat sich die Audiotechnik an die rasante Entwicklung der digitalen Speichermedien angehängt: Aktuell sind CD-ROMs, Mini-Disks und MPEG3-Player.

laufend neue Abspielgeräte!

Wer von diesem Fortschritt profitieren möchte, sieht sich genötigt, jeweils nach einigen Jahren seine Datenträger (und Abspielgeräte!) komplett auszutauschen: Schellacks gegen Langspielplatten, Mono- gegen Stereoplatten, Platten und Bänder nacheinander gegen CD-ROMs und Minidisks. Dies nützt der Industrie und beflügelt den Fortschritt. Wer nicht mitmacht, steht bald ohne Unterstützung da, weil neue Tondokumente nicht in veralteter Technologie angeboten werden und weil für die Abspielgeräte keine Ersatzteile zu bekommen sind. Schellackliebhaber finden Platten und passende Grammophone allenfalls beim Trödler oder im Mu-

seum. Auch wenn die neue Technologie prinzipiell das Kopieren (und dadurch Erhalten) der alten Tondokumente erlaubt, wird wegen Mängeln der alten Träger (Rauschen, fehlende Stereophonie etc.) vielfach darauf verzichtet.

Völlig analog, nur wegen der erheblich kürzeren Innovationszyklen wesentlich dramatischer verläuft die Entwicklung in der Digitaltechnik.

Digitaltechnik noch schnelllebig

Versetzen wir uns nun in die Rolle einer Musikwissenschaftlerin, die Originalaufnahmen des großen Enrico Caruso studieren möchte. Je nach Zweck der Untersuchung wird sie einen von zwei grundverschiedenen Zugängen wählen. Zum einen kann sie versuchen, zeitgenössische (Schellack-)Aufnahmen auf einem passenden Grammophon abzuspielen. Zum anderen kann sie sich eine mit Rauschfiltern und anderen technischen Transformationen nachträglich auf den heutigen Stand gebrachte Aufnahme beschaffen. Wenn es darum geht, die Interpretationen einer bestimmten Arie durch Caruso und Pavarotti zu vergleichen, ist der zweite Weg vermutlich angemessen. Soll dagegen ein größtmögliches Maß an Authentizität sichergestellt werden, dann wird unsere Musikwissenschaftlerin eher den ersten Weg beschreiten.

Authentizität einer Aufnahme

Technische Maßnahmen zur elektronischen Langzeitarchivierung lassen sich in die gleichen Kategorien einteilen: auf der einen Seite Maßnahmen, die möglichst nah am Originaldokument und der Originalabspielumgebung zu bleiben trachten; auf der anderen Seite Maßnahmen, welche das Dokument fortlaufend dem neuesten Stand der Technik anzupassen versuchen.

Originale erhalten oder anpassen?

Migration

In der Datenverarbeitung wie in der Audiotechnik werden Maßnahmen der zweiten Kategorie häufig und auf breiter Front eingesetzt. Da hierbei Daten fortlaufend von einer Darstellungsform zur nächsten »wandern«, spricht man von *Migration*. Es bietet sich an, diese bekannten und bewährten Verfahren auch zur Langzeitarchivierung elektronischer Dokumente zu verwenden. Die Migration von Zeichenströmen wurde in Abschnitt 1.3 kurz erwähnt. Die Vorteile der Migration von Dokumenten liegen auf der Hand:

Migration

Vorteile der Migration

- Da Migration in der Datenverarbeitung allgemein angewandt wird, kann man auf Erfahrungen, Methoden, geschultes Personal und vor allem auf vorhandene Werkzeuge zurückgreifen.

allgemein bekannt

stets aktuell

- ❑ Migrierte Dokumente sind prinzipiell zu jedem Zeitpunkt auf den gerade aktuellen Abspielsystemen verfügbar. Sofern dem nicht juristische oder organisatorische Gründe entgegenstehen, ist ein weltweiter Zugriff in Sekundenschnelle kein Problem.

eventuell

Qualitätsverbesserung

- ❑ Auch die Qualität migrierter Dokumente entspricht den aktuellen Standards: So wie bei Tonaufzeichnungen heute Stereophonie und HiFi-Qualität selbstverständlich sind, werden bei Textdokumenten moderne Fonts und ein gefälliges Layout erwartet.

Nachteile der Migration

Der letzte Punkt deutet aber auch schon die Nachteile der Migration an:

Verfälschungen des Originals

- ❑ Die bei der Migration notwendigen Anpassungen führen fast immer zwangsläufig zu Verfälschungen des Originals, d. h., die Authentizität des Dokuments ist nicht gewährleistet. Ein anderer Bereich, in dem über lange Zeit hin mit Migration gearbeitet wurde, ist die Bibelübersetzung. Hier wird mühsam auf frühere Fassungen zurückgegriffen, um die Authentizität zu verbessern.

schwer automatisierbar

- ❑ Die bei der Migration anfallenden Transformationen sind nur in einfachen Fällen automatisch (per Programm) durchführbar. Ein solcher gut beherrschbarer Schritt wäre z. B. die Umwandlung von ASCII-Dokumenten in Unicode-Dokumente (wobei das Erkennen von Ersatzdarstellungen von in Unicode zusätzlichen Zeichen ein manuelles Nachbearbeiten erforderlich machen kann). Wie in Abschnitt 1.2 am Beispiel der Liniengraphik angedeutet, können im alten Layout Informationen implizit enthalten sein, die sich nur dem Auge des verständigen Lesers erschließen.

Die Migration ist Gegenstand von Kapitel 3.

Hardware-Museen

Hardware-Museen sind nicht praktikabel

Die Maßnahme zur elektronischen Langzeitarchivierung, die am stärksten auf den Erhalt des Originalzustands ausgerichtet ist, besteht in der Einrichtung so genannter *Hardware-Museen*. Die Aufgabe von solchen Hardware-Museen ist es, möglichst lückenlos alle bekannten Computer betriebsbereit zu erhalten, damit künftige Generationen unsere elektronischen Dokumente in ihrer ursprünglichen Umgebung betrachten können. An Authentizität ist



Abbildung 1.3
*Hardware-Museum an
der Universität der
Bundeswehr München*



dieser Ansatz nicht zu übertreffen. Leider erweist er sich aus einer Reihe von Gründen als nicht praktikabel:

- ❑ Selbst heute schon erscheint es unrealistisch, sämtliche bisher entwickelten Computertypen in einem Museum zu ver-

zu viele Exponate

unterschiedlichste
Software und
Hardwarezusätze

begrenzte
Lebensdauer

- sammeln, geschweige denn dies für die Zukunft sicherzustellen.
- ❑ Wie in Abschnitt 1.2 ausgeführt, besteht eine vollständige Abspielumgebung nicht nur aus Hardware, sondern vor allem auch aus System- und Präsentationssoftware. Neben der Hardware müsste das Museum also auch sämtliche Versionen von Betriebssystemen und Anwendungsprogrammen für ihre Computer vorhalten und bei Bedarf installieren. Gleiches gilt für spezielle Hardwarezusätze (z. B. Joysticks oder Digitalisiertabletts) und ihre Treibersoftware.
 - ❑ Auch technische Geräte und ihre Bausteine haben nur eine begrenzte Lebensdauer. Da weder die Geräte noch ihre Bausteine weiter produziert werden, müsste man genaue technische Spezifikationen zu Aufbau und Funktionsweise des Gesamtsystems zusammen mit den technischen Leistungsmerkmalen der Komponenten aufbewahren, nach denen die Wartungstechniker bei Bedarf Ersatzteile herstellen könnten (sofern die zur Fertigung benötigten Techniken noch bekannt sind).

Ein solches Projekt wäre wegen seines enormen Bedarfs an materiellen und personellen Ressourcen zum Scheitern verurteilt.

Emulation

Emulatoren in der
Hardwareentwicklung

Bei der Entwicklung neuer Hardware werden – um Zeit und Kosten zu sparen – so genannte Emulatoren eingesetzt, mit denen man die Funktionsweise von Prozessoren und anderen Hardwarekomponenten durch Software simulieren und testen kann, bevor man sie tatsächlich baut. Von Rothenberg (1999) stammt der Vorschlag, *Emulation* zur elektronischen Langzeitarchivierung zu verwenden.

Vorgehen beim
Emulationsansatz

Beim Emulationsansatz werden von jedem elektronischen Dokument nicht nur der Zeichenstrom und die Metadaten konserviert (z. B. durch Migration), sondern zusätzlich eine Emulatorgeeignete Spezifikation der Hardware und binäre Zeichenströme, welche die komplette Software der Abspielumgebung enthalten. Will jemand in beispielsweise 100 Jahren ein so konserviertes Dokument betrachten, dann ist in folgenden Schritten vorzugehen:

Emulator herstellen,
...

1. Verwende die Spezifikation der Hardware, um zusammen mit dem Emulator eine Software-Implementierung der Originalhardware herzustellen, die sich funktional identisch verhält. (Dies ist der entscheidende Schritt und setzt voraus, dass die

Hardwarespezifikation in einem vom Emulator verarbeitbaren Format vorliegt.)

2. Installiere auf dem emulierten Computer die Systemsoftware und die benötigten Präsentationsprogramme. (Die Informationen zur Handhabung dieses 100 Jahre alten Computertyps, insbesondere zur Installation von Software, sollten daher zusammen mit dem Dokument abgelegt werden.)
3. Lade den Zeichenstrom des elektronischen Dokuments in den emulierten Computer und starte das passende Präsentationsprogramm. (Hierzu empfiehlt sich die Speicherung des Handbuchs zum Präsentationsprogramm mit dem Dokument.)

... darauf Software installieren ...

... und Dokumente »abspielen«

Der Emulationsansatz wird detailliert in Kapitel 4 beschrieben.

Trotz seiner offenkundigen Komplexität ist die Langzeitarchivierung mit Emulation aus einer Reihe von Gründen attraktiv:

Vorteile der Emulation

- ❑ Ist ein elektronisches Dokument erst einmal in die für die Emulation benötigte Form gebracht, dann sind später keine weiteren Verarbeitungsschritte mehr nötig (wenn man von den Maßnahmen zur Konservierung des zugrunde liegenden Zeichenstroms und der Migration des Emulators einmal absieht). Im Unterschied dazu erfordert Migration in regelmäßigen Abständen Anpassungen an einen immer anwachsenden Dokumentenbestand, wobei die Anpassungen auch nur teilweise zu automatisieren sind.
- ❑ Da wie beim Museumsansatz mit den Datenströmen der Originaldokumente gearbeitet wird, beschränkt sich der bei der Konservierung anfallende Aufwand im Wesentlichen auf das Erstellen der Hardwarespezifikationen, das aber nicht bei jedem Dokument, sondern »nur« bei jedem neuen Computertyp zu leisten ist. Wünschenswert wäre eine Übernahme der Emulatorspezifikationen der Hardwarehersteller aus der Entwicklungsphase.
- ❑ Die Authentizität der konservierten Dokumente ist ähnlich hoch wie beim Museumsansatz, da das Dokument praktisch nicht verändert wird.
- ❑ Der mit der Emulation verbundene laufende Aufwand besteht hauptsächlich aus den – allerdings hohen – Zugriffskosten, ist also proportional zur Verwendungshäufigkeit.

Aufwand je Dokument gering

»nur« ein Emulator für viele Dokumente

hohe Authentizität

Aufwand proportional zur Verwendungshäufigkeit

Die Emulation befindet sich noch in der Erprobungsphase. Während man bei der Migration jeweils kleine, überschaubare Schritte

Probleme bei der Emulation

macht und Lösungen immer dann gefunden werden müssen, wenn sich die Computertechnologie ein Stück weiterentwickelt, hängt die Durchführbarkeit des Emulationsansatzes von positiven Antworten auf folgende Fragen ab:

- | | |
|---|---|
| <i>präzise Beschreibung der Hardware</i> | <ul style="list-style-type: none"> ❑ Findet man eine Hardware-Beschreibungssprache, in der man die Spezifikationen eines breiten Spektrums von Hardware mit ausreichender Genauigkeit fassen kann? Damit künftige Entwicklungen berücksichtigt werden können, sollte die Sprache erweiterbar sein. Alle Eigenschaften, welche an der Benutzerschnittstelle sichtbar sind (Bildschirmgröße, -auflösung und -farben, Taktrate des Prozessors etc.) müssen mit aufgeführt werden. |
| <i>Generierung von Emulationsprogrammen</i> | <ul style="list-style-type: none"> ❑ Lassen sich aus solchen Beschreibungen automatisch Emulationsprogramme generieren, die laufzeiteffizient genug sind, um die Realzeit-Eigenschaften der nachzubildenden Hardware (z. B. wenn Musik oder Filmsequenzen wiedergegeben werden sollen) zu erfüllen? |
| <i>portable Emulatoren?</i> | <ul style="list-style-type: none"> ❑ Kann der Emulator ausreichend portabel gestaltet werden? Theoretisch könnte man ein dezidiertes Emulatorsystem (als Hardware) bauen. Sinnvoller erscheint es aber, eine Emulatorsoftware herzustellen und jeweils auf alle wichtigen Plattformen zu portieren, damit der Emulator von den technologischen Fortschritten profitieren kann und auch überall verfügbar ist. |
| <i>Speicherungsredundanz vermeiden</i> | <ul style="list-style-type: none"> ❑ Lässt sich eine praktikable Organisation finden, welche der oben beschriebenen Speichertechnik inhaltlich gleichwertig ist, aber deren Redundanz (Computerspezifikation und Software bei jedem Dokument) vermeidet, ohne die Integrität der Dokumente zu gefährden? |
| <i>Aufwand wegen Vielzahl der Formate</i> | <p>Ganz besonders bei der Migration, aber auch bei der Emulation gilt: der für die Langzeitspeicherung zu treibende Aufwand steigt bzw. fällt mit der Anzahl der zugelassenen Datenformate. Es liegt daher nahe, Vielfalt und Wildwuchs durch Einführung von Standards zu beschneiden, um den Aufwand in einem handhabbaren Rahmen zu halten.</p> |
| <i>Standards für einfache Zeichenfolgen</i> | <p>Die in Abschnitt 1.3 erwähnten Standards (ASCII, Unicode) für einfache Zeichenfolgen sind wohletabliert und können auf absehbare Zeit als stabil gelten. Das macht sie besonders attraktiv als Grundlage für die Speicherung der Metadaten, die den Inhalt von elektronischen Dokumenten und deren Handhabung beschrei-</p> |

ben. Es ist wichtig, diese Zugriffs- und Zugangsschlüssel in einem einfachen und verlässlichen Format aufzubewahren.

Je komplexer und gebräuchlicher die Dokumentenarten sind, desto größer die Palette an unterschiedlichen, konkurrierenden Formaten: Für Textdokumente mit eingebetteten Graphiken reicht das Spektrum von internen, proprietären Formaten (z. B. Word) über prozedurale Beschreibungen (PostScript) und Anzeigeformate (DVI und PDF) bis hin zu Auszeichnungssprachen (L^AT_EX und XSL). Auf der Basis von XML entsteht derzeit eine Reihe von »selbstbeschreibenden« Formaten, die für künftige Standardisierungsbemühungen möglicherweise richtungweisend sein werden. Dazu gehören RDF und Dublin Core für Metadaten, XSL für Desktop Publishing, SVG für Vektorgraphiken, SMIL für Multi-Mediadokumente und MathML für mathematische Formeln.

*Standards für
komplexe
Dokumentenarten*

Auszeichnungssprachen auf der Basis von SGML und XML werden daher den Schwerpunkt des Kapitels 6 über Standardformate bilden.

Auszeichnungssprachen

1.5 Juristische und gesellschaftliche Fragen

Mit den von uns behandelten technischen Aspekten der elektronischen Langzeitarchivierung sind juristische und gesellschaftliche Fragestellungen verknüpft, die in den folgenden sieben Thesen nur kurz angerissen werden.

sieben Thesen

1. Bevor elektronische Dokumente archiviert werden können, sind *Copyright*-Fragen zu klären: Da insbesondere bei der Migration von Dokumenten Kopien erstellt werden und evtl. Originaldokumente verändert werden müssen, ist das Einverständnis des Urhebers mit solchen Maßnahmen vorab einzuholen. Weitere Kopien, die an Leser von Dokumenten ausgehändigt werden, sind angemessen zu honorieren und müssen gegebenenfalls mit Sperrvermerken verbunden werden, wenn eine unentgeltliche Weitergabe nicht erlaubt ist. Der Unterschied zwischen herkömmlichen und elektronischen Dokumenten ergibt sich daraus, dass bei letzteren Kopie und Original praktisch nicht zu unterscheiden sind. Basis einer für alle Seiten handhabbaren Lösung könnte eine entsprechende Ausweitung der GEMA-Regelungen sein.
2. Beim Emulationsansatz sind (mittelfristig) zusätzlich die Rechte an den Software-Anteilen der Originalabspielumge-

*»Digital Rights
Management« und
Kopierschutz*

*Rechte an der
Software*

- Emulatoren von den Hardwareherstellern?
- Kriterien zur Selektion
- Kosten als begrenzender Faktor
- bung zu wahren. Auch hier wäre eine pauschale, zweckgebundene Regelung nützlich.
3. Soll der Emulationsansatz Anwendung finden, wäre es sinnvoll, die Hardwarehersteller gesetzlich zu verpflichten, nach einer angemessenen Frist vollständige, für Emulationszwecke geeignete Spezifikationen ihrer Geräte der Allgemeinheit zur Verfügung zu stellen. Da aktuelle Geräte verfügbar sind, müsste eine solche Verpflichtung erst dann greifen, wenn die Entwickler den vollen wirtschaftlichen Nutzen aus dem eher kurzlebigen Innovationsgehalt ihrer Produkte gezogen hätten.
 4. Arbeitsgrundlage für Bibliotheken und Archive ist eine klare Aussage darüber, was eigentlich als bewahrenswert zu betrachten ist. Für Druckerzeugnisse gibt es Kriterien (z. B. die Existenz einer zugehörigen ISBN-Nummer) und gesetzlich abgesicherte Verfahrensweisen (Verlage müssen Belegexemplare liefern). Es wäre zu kurz gegriffen, nur solche elektronischen Dokumente als relevant zu betrachten, die an greifbare Datenträger wie CD-ROMs gebunden und mit einer ISBN-Nummer versehen sind. Wissenschaftliche elektronische Zeitschriften und weitere, »nur« im Internet existierende Publikationen gehören sicherlich auch zum bewahrenswerten kulturellen Erbe. Neue Kriterien müssen aufgestellt und vom Gesetzgeber durch geeignete Verfahrensweisen unterstützt werden. Wegen der Fülle des Materials und der hohen Kosten sind insbesondere *Kriterien zur Selektion* wichtig und sollten sorgfältig überdacht werden. Andererseits drängt die Zeit, da jetzt schon Abspielumgebungen laufend veralten. Anders als bei Druckerzeugnissen kann man nicht gut abwarten, was langfristig nachgefragt wird und dadurch seinen Wert belegt.
 5. *Kosten* sind ein entscheidender begrenzender Faktor. Auf nationaler und internationaler Ebene sind langfristig tragende Vereinbarungen erforderlich, welche die Träger der elektronischen Langzeitarchivierung sowie ihre Zuständigkeiten und finanzielle Ausstattung verlässlich festlegen. Natürlich sind solche Vereinbarungen laufend fortzuschreiben und damit nicht endgültig fixiert; ein tragfähiges Rahmenprogramm ist aber Grundvoraussetzung für eine dauerhafte Kulturarbeit. Sinnvoll ist es sicher auch, durch Kooperation und Aufgabenteilung global Kosten zu senken: Gemeinsame Standards sind ein Aspekt, Verteilung von einander ergänzenden Schwerpunkten ein anderer.

6. Kosten können allein von staatlicher Seite getragen oder zum Teil von anderen übernommen werden, z. B. die Zugriffskosten in Form von Lesegebühren und die Kosten der erstmaligen Archivierung (nebst entsprechender Datenaufbereitung) von Autoren, die ihre Produkte erhalten sehen möchten. Mit Vorschriften über die Form der eingereichten Dokumente (zulässige Standardformate, erforderliche Metadaten) ließen sich die Kosten der Archivierung ebenfalls senken. *Kosten tragbar machen*
7. Alle Regelungen müssen die berechtigten Interessen der beteiligten Parteien (Autoren und Verlage, Bibliotheken und Archive, Wissenschaftler und Öffentlichkeit) ausgewogen berücksichtigen. Die verschiedenen Standpunkte werden von Haynes *et al.* (1997) ausführlich einander gegenübergestellt. *Interessenausgleich zwischen allen Beteiligten*

Einige dieser Fragen werden wir in der Diskussion in Kapitel 7 nochmals aufgreifen.