

# Herausforderungen an die digitale Erhaltung

HANS HOFMAN

## 1 Einführung

Der Einfluß der Informationstechnologie auf die Gesellschaft, im Büro und auch zu Hause ist bekannt. Die Informationstechnologie hat unsere Lebensweise tiefgreifend verändert oder wird das bald tun. Der Begriff *Informationsgesellschaft* ist sehr häufig zu hören, aber was er eigentlich bedeutet, ist unklar. Ich weiß nicht, ob wir diese scheinbar vielversprechende Zukunft schon erreicht haben. Dennoch sind wir Zeugen dessen, was sich alles sehr schnell unter dem Einfluß dieser neuen Technologie verändert und wir fragen uns, wie wir uns darauf einstellen sollen oder vielleicht müssen.

Eine der Fragen, die häufig diskutiert wird, ist die nach dem Zugang zu Informationen oder, um es präziser zu formulieren, nach dem Zugang zur kulturellen Überlieferung. Der Zugang zu öffentlichen Datenbeständen, und diese sind Teil der kulturellen Überlieferung, wird in demokratischen Gesellschaften als eines der Grundrechte angesehen. Daß dies nicht immer ausreichend oder befriedigend geregelt ist und daß jeder Mitgliedstaat der Europäischen Union es anders geregelt hat, ist eine andere Sache. Die Bibliotheken, Museen und Archive haben auf dem Gebiet der Aufbewahrung kultureller Überlieferung eine Schlüsselstellung. Die Artefakte und Objekte, die sie aufbewahren, sind derzeit meistens noch aus Holz, Stein, Papier oder aus anderen traditionellen Materialien gemacht. Um diese Objekte im Internet oder in digitalen Bibliotheken zur Verfügung zu stellen, muß man sie digitalisieren. Dennoch hat man immer noch die ursprünglichen physischen Gegenstände.

Die Realität ist, daß es mit der wachsenden Anwendung der Informationstechnik immer mehr Information in digitaler Form gibt, nicht nur in Behörden und Bibliotheken, sondern auch in der Wirtschaft. Ich nenne hier zum Beispiel elektronische Publikationen, wissenschaftliche Dateien und digitale Akten. Diese Informationen sind unmittelbar in digitaler Form erzeugt. Der große Unterschied zu traditionellen Artefakten und Informationsträgern ist, daß diese unmittelbar wahrnehmbar und greifbar sind. Digitale Objekte sind dagegen nicht unmittelbar wahrnehmbar, nicht greifbar und außerdem sehr flüchtig.

Eine der wirklich großen Herausforderungen in der sogenannten Informationsgesellschaft ist daher auch die Frage, wie diese neuen „Kulturäußerungen“ aufbewahrt werden können. Viele unserer frühen Vorfahren konnten nicht schreiben, und daher haben wir zum Beispiel wenig Informationen über das Leben der einfachen Leute im frühen Mittelalter. Das Paradoxon, mit dem wir es im Vergleich zu unseren Vorfahren zu tun haben, ist, daß wir heutzutage immer mehr Daten produzieren, aber eigentlich nicht imstande sind, diese gut aufzubewahren. Ist das die Informationsgesellschaft, die wir wollen?

Die Herausforderung ist also deutlich: Wir müssen jetzt die notwendigen Maßnahmen zur digitalen Aufbewahrung ergreifen, um unser kulturelles Erbe, einschließlich der „kulturellen Äußerungen“, die heute erzeugt werden, für künftige Generationen zu sichern.

Ich möchte in meinem Vortrag deutlich machen, welche wichtigen und wesentlichen Veränderungen es gibt und was die zentralen Themen auf diesem Gebiet sind. Eine Botschaft soll dabei von vornherein deutlich sein: Archive und Bibliotheken können das Problem der langfristigen Aufbewahrung der digitalen Überlieferung nicht alleine lösen. Sie benötigen dazu die Unterstützung von Wirtschaft und Wissenschaft.

## 2 Änderungen und Themen

Obwohl Sie vielleicht wissen, welche Entwicklungen es in der Informationstechnologie gibt und was die Konsequenzen für die Aufbewahrung von Registratur- und Archivgut, elektronischen Publikationen und digitalen Informationen im allgemeinen sind, möchte ich doch einen kurzen Überblick über einige dieser Entwicklungen geben, die insbesondere für die Erhaltung digitaler Bestände entscheidend sind.

Die wichtigste Konsequenz der IT-Anwendung bei der Erzeugung digitaler Informationen ist die Änderung der Natur des Dokuments. Wie schon gesagt, das Dokument als physische Entität verschwindet, und das ist einer der interessantesten Aspekte des digitalen Zeitalters. Die vertrauten Akten und Veröffentlichungen aus Papier werden durch abstrakte und dem Wesen nach unsichtbare digitale Dokumente und Akten ersetzt. Unterlagen aus Papier und Pergament sind physische Vertreter einer vergangenen Welt und erzählen uns über längst vergangene Aktivitäten. An sich sind sie auch Artefakte, die uns einen Eindruck geben, wie es einmal war, einfach weil man sie berühren und riechen kann. Während digitale Dokumente demselben Ziel dienen, sind sie aber nicht unmittelbar wahrnehmbar. Wenn man nicht das richtige Instrument hat, kann man sie nicht sehen oder benutzen, geschweige denn berühren oder riechen. Das verlangt vom Bibliothekar oder Archivar einen echten Paradigmenwandel im Denken.

Das Dokument ist durch die Umwandlung in die digitale Form in verschiedene Komponenten auseinandergefallen. Diese Komponenten sind:

- ein digitaler Bestand mit verschlüsselten Daten,
- ein Träger, auf dem der Bestand gespeichert ist,
- Hardware,
- ein Programm
- und Informationen, die notwendig sind um zu wissen, wie man den digitalen Bestand lesen und zur Verfügung stellen kann sowie um zu verstehen, worum es in den Dokumenten geht.

Dazu kann man auch noch Kontext-Informationen rechnen, die uns über die Provenienz und die Rolle informieren, die das Dokument (oder die Akten) in einem Geschäftsvorgang einmal gespielt hat, um eine Aufgabe oder Funktion zu erfüllen.

Wenn man das überblickt, dann entdeckt man logische und physische, oder besser technische Aspekte. Das Dokument ist nur dann vorhanden, wenn es auf dem Bildschirm präsentiert wird, und um das zu erreichen, sind die genannten fünf Komponenten nötig. Man kann es auch anders und deutlicher sagen: Ohne all diese Komponenten existiert ein digitales Dokument nicht.

Diese neue Situation hat einen großen Einfluß auf die Aufbewahrung sowie auch auf die Verwaltung digitaler Informationen. Eine der Konsequenzen ist, daß wir einen Unterschied zwischen dem logischen oder intellektuellen Aspekt und dem technischen Aspekt machen müssen. Jeder dieser Aspekte hat seine eigenen Objekte, die verwaltet und erhalten werden müssen.

- Die intellektuelle Ebene betrifft die Darstellung digitaler Informationen auf eine Weise, daß Menschen sie verstehen und benutzen können. Das ist eigentlich das Ziel unserer Aufgabe.
- Auf der technischen Ebene ist das erste Ziel die Erhaltung von digitalen Daten. Es umfaßt alle Aktivitäten, die notwendig sind, um digitale Informationen zu reproduzieren und zu präsentieren, so daß Menschen sie lesen können. Die Technologie ist also notwendig, um Akten und andere Informationen dauerhaft und trotz des raschen technologischen Wandels zu bewahren.

Dies ist eine echte Herausforderung. Die Geschichte der Informationstechnologie ist eine Geschichte

sehr schneller Entwicklungen, und dabei scheint der ständige Wandel der normale Zustand zu sein. Außerdem ist klar, daß das Ende dieser Entwicklung nicht kurzfristig zu erwarten ist, wenn überhaupt. Vielleicht wird sich die Geschwindigkeit der Entwicklung nach einiger Zeit verringern. Was die Aufgabe wirklich schwer und kompliziert macht, ist die Tatsache, daß jede der technischen Komponenten ihre eigene Entwicklungsgeschwindigkeit und ihren eigenen Lebenszyklus hat.

Die neueren Entwicklungen in der Informationstechnologie, wie die Anwendungen von Multimedia oder komplexen Datenbanken, ermöglichen ganz neue Typen von Dokumenten oder Informationsarten, die nicht nur Texte, sondern auch Bilder, Grafiken, Spreadsheets und akustische Sequenzen und Video umfassen und daher nicht auf Papier ausgedruckt werden können.

Diese neuen Typen von Dokumenten verlangen auch einen neuen Archivierungsansatz, damit sie nicht verloren gehen. Tatsächlich werden sie allerdings meistens noch nicht als Informationen, die aufbewahrt werden müssen, angesehen. Außerdem bedeutet dies, daß die traditionelle Form oder Erscheinung eines Dokumentes sich ändert. So verschwindet zum Beispiel die physische Beschränkung auf eine DIN-A-4-Seite Papier in der digitalen Welt.

Alle diese Entwicklungen führen dazu, daß die heutigen Prozeduren und Verfahren nicht länger ausreichend sind und daß sie revidiert werden müssen. Dieser Änderungsbedarf gilt sicherlich für spezialisierte Einrichtungen wie Archive und Bibliotheken. Er gilt jedoch auch für Behörden und die Wirtschaft, zum Beispiel für die pharmazeutische Industrie, Bohrbetriebe, Ölgesellschaften etc. Anderenfalls werden wir den Zugriff auf Dokumente verlieren, gerade weil dies ein virtuelles Phänomen (geworden) ist.

Mit dieser Schlußfolgerung kommt die nächste und wichtigste Frage, wenn wir digitale Informationen, wie Akten, Veröffentlichungen, Forschungs-Datenbestände etc. aufbewahren wollen: Mit welchen Anforderungen haben wir es dann zu tun und wie sollen wir diese erfüllen?

Ich möchte die drei wichtigsten Anforderungen nennen:

1. *Authentizität*: Wie bei Papier-Akten ist es wichtig, sehen zu können, ob ein Dokument ursprünglich und authentisch ist. Es soll somit möglich sein, festzustellen, ob das Dokument noch dasselbe ist wie zu dem Zeitpunkt, als es im Geschäftsgang war. Nur dann bieten Dokumente am besten einen Nachweis darüber, was geschehen ist, und beweisen, daß sie das sind, was sie zu sein vorgeben. Die Trennung zwischen dem physischen und dem logischen Aspekt bei digitalen Dokumenten erfordert eine höhere Aufmerksamkeit.
2. *Dauerhaftigkeit*: Die Flüchtigkeit der Einsen und Nullen und die schnellen Änderungen in der Informationstechnologie erfordern spezielle Maßnahmen auf dem Gebiet der Aufbewahrung. Sie sollen es möglich machen, daß auch unsere Ur-Ur-Enkel noch sehen können, wie die Regierung in unserer Zeit gehandelt hat.
3. *Verständlichkeit*: Wir können vielleicht alles gut und authentisch aufbewahren, aber wenn es nicht möglich ist, diese Informationen zu verstehen, dann haben zukünftige Forscher nichts davon. Die Akten sollen also verständlich und mit ausreichenden Kontext-Informationen versehen sein, die erklären, wo und wann sie entstanden sind und welche Rolle sie bei welchem Geschäftsvorgang gespielt haben.

Wenn wir versuchen, die richtigen Fragen zur Aufbewahrung zu identifizieren, die auf diesen Anforderungen gegründet sind, und einen Blick auf die technischen Komponenten zu werfen, dann entdecken wir, daß insbesondere das Speicherformat und das Anwendungsprogramm wichtig sind. Diese Komponenten bestimmen im Grunde die Authentizität der Akten. Wenn sie sich aufgrund des technologischen Wandels ändern, dann werden sich wahrscheinlich auch das Aussehen und die Struktur verändern. Um ein einfaches Beispiel zu geben: Sogar wenn ein Dokument innerhalb eines Textverar-

beitungspakets wie MS WORD zu einer neuen Version konvertiert wird, dann können die Struktur (und der Bitbestand) dieses Dokuments beeinflußt werden mit der Folge, daß die Authentizität und die Beweiskraft eingeschränkt werden.

Wenn wir diese Fragen und den Wandel der Anforderungen betrachten, so stellt sich die Frage, wie wir damit umgehen können. Sind da schon Lösungen oder mögliche, vielversprechende Verfahren vorhanden?

### 3 Mögliche Lösungsrichtungen

Migration wird heute gewöhnlich als die effektivste Lösungsrichtung angesehen. Dahinter steht der Gedanke, daß die digitalen Informationen durch regelmäßige Konversionen in neue Formate durch die Zeit und den technologischen Wandel hindurch bewahrt werden können. Migration ist meistens abhängig von einer Standardisierung der Speicherformate. Standardisierung ist auch eine Methode, um die Verwaltung von digitalen Informationen beherrschbar zu machen. Die Frage ist jedoch, ob dies der richtige Ansatz sein wird. Einer der Nachteile der Konvertierung und Migration ist, daß bei jeder Konvertierung oder Migration Daten verloren gehen. Und das ist es gerade, was wir vermeiden möchten. Unser Ziel ist doch, digitale Dokumente (Akten) in ihrer ursprünglichen und authentischen Form aufzubewahren.

Auf welche Weise können wir dieses Problem lösen? Gibt es andere, geeignetere und zweckmäßigere Annäherungen? Das ist bislang noch offen. In den Niederlanden haben wir ein Projekt angefangen mit dem Ziel, ein besseres Verfahren durch die Entwicklung einer Testumgebung („test-bed“) zu entwickeln, in der experimentiert werden kann.

Wir haben daher RAND Europe gebeten, einen Bericht zu verfassen, einen ersten Entwurf zu erstellen und weitere Empfehlungen über die Untersuchungsfragen und deren Fortsetzung zu geben. Der damit beauftragte Forscher ist Jeff Rothenberg, ein bekannter Computerwissenschaftler, der ein starker Befürworter der Emulation ist.

Für die genannten Anforderungen, wie beispielsweise Authentizität, erscheint dieser Lösungsansatz der vielversprechendste zu sein.

Aufgrund der Kürze der Zeit kann ich nur die wichtigsten Merkmale dieses Lösungsansatzes nennen, diese sind:

- Die Aufbewahrung der authentischen digitalen Akten mit der ursprünglichen Software (das Ziel),
- um das zu erreichen, wird die alte Hardwareumgebung auf künftigen, jetzt noch unbekanntem Systemen emuliert,
- und das ursprüngliche digitale Dokument wird zusammen mit der ursprünglichen Software-Umgebung (sowohl das Anwendungsprogramm als auch die System-Software) sowie mit einer Spezifizierung, einem „Emulator“ für die ursprüngliche Computerplattform des Dokuments und mit erklärendem Material und anderen Informationen „eingekapselt“ (encapsulated).

Was emuliert werden muß, ist vor allem der Prozessor, auf welchem die Software läuft. Die Vorteile dieser Annäherung sind klar:

- Für jede Generation der Hardware, oder besser jeden Prozessortyp braucht man nur einmal ein Emulationsprogramm zu schreiben,
- dieses Programm ist einheitlich und automatisch für alle Dokumente (Akten) anwendbar und daher sehr zweckmäßig,
- mit dieser Methode wird auch die Authentizität und die ursprüngliche Funktionalität des Dokuments erhalten.

Allerdings sind noch viele Probleme vorhanden:

- Es gibt noch keine standardisierte Beschreibungsweise für Emulatoren,
- die Dokumentation, beispielsweise von Software-Programmen, ist meistens unzureichend, wenn sie überhaupt existiert.

Obwohl noch überprüft werden muß, wie dauerhaft und geeignet dieser Lösungsansatz ist, erscheint er mir trotzdem als ein sehr interessantes Verfahren, das sicher wert ist, näher untersucht und erprobt zu werden.

Ob dies die einzig dauerhafte Lösungsrichtung ist oder nicht, das ist eine Frage, die wir in der Zukunft beantworten können.

#### **4 Zusammenfassung**

1. Eine der wichtigsten Fragen in der Informationsgesellschaft, die leicht vergessen wird, weil Informationen immer überall anwesend zu sein scheinen, ist die nach der langfristigen Erhaltung von digitalen Unterlagen und Akten in ihrer ursprünglichen (authentischen) Form. Wenn wir im digitalen Zeitalter noch etwas für künftige Generationen hinterlassen wollen, dann sollten, nein müssen wir jetzt etwas tun.
2. Die Einrichtungen, die an erster Stelle mit dieser Herausforderung zu tun haben, sind die Archive und Bibliotheken. Das ist hauptsächlich auf ihre Aufgabe oder ihr Mandat zurückzuführen, das sie für die Erhaltung von Informationen (Akten oder Publikationen) verantwortlich macht, die sie in einer authentischen Form zu verwalten haben. Das erscheint beinahe als eine unmögliche Aufgabe in dieser sich ständig ändernden Welt der Informationstechnik. Trotzdem ist es eine Voraussetzung für die Informationsgesellschaft und den Zugang zum kulturellen Erbe, um das gesellschaftliche Gedächtnis für die Zukunft zu erhalten.
3. Es gibt leider noch keine endgültigen Lösungen, nur Ideen und Möglichkeiten. Eine dieser Möglichkeiten ist Emulation, wie sie von Jeff Rothenberg vorgeschlagen wurde. Sie scheint auch vielversprechend zu sein, bedarf aber noch der Erprobung in der Praxis. Soweit bekannt, ist das niederländische Projekt eines der ersten, das eine praktische Umsetzung des Konzepts versucht.
4. Last not least fordert diese Aufgabe auch Zusammenarbeit, um die Lösungen zu finden. Dies sollten an erster Stelle multidisziplinäre, aber eigentlich auch internationale Projekte sein.

Bis jetzt werden die heutigen Projekte hauptsächlich durch kulturelle Einrichtungen ausgeführt oder unterstützt, beispielsweise das europäische *NEDLIB Project* oder das niederländische *digitale Depot*. Aber auch Wirtschaftsunternehmen sollten mit einbezogen werden – und manchmal werden sie das auch –, so beispielsweise die pharmazeutische Industrie, weil sie dasselbe Problem der langfristigen Aufbewahrung von Daten in authentischer Form haben. Die IT-Industrie ist ein anderer potentieller Partner, der aber bis jetzt noch nicht einbezogen wurde. Das sind jedoch ganz andere Herausforderungen!