## Rematerialisierung digitaler Unterlagen – ein Weg zur Langzeitarchivierung?

## Das Projekt HD-Rosetta

Von Karl-Ernst Lupprian

Es hieße Eulen nach Athen tragen, wenn man in diesem Kreis die Faktoren aufzählen wollte, die das Überleben von Unterlagen in digitaler Form gefährden. Charles Dollar hat sie in seinem jüngsten Buch über authentische elektronische Unterlagen¹ ausführlich diskutiert und Strategien für eine Langzeitaufbewahrung vorgeschlagen. Dabei favorisiert er mehr oder weniger die Migration, also das rechtzeitige Konvertieren der Daten in Formate, die auf jeweils neuen Hard- und Softwareplattformen verarbeitet werden können.

Zu Recht warnt Dollar vor den im voraus kaum kalkulierbaren, aber mit Sicherheit hohen Kosten für Migrationen. Schon die Aufbewahrung digitaler Unterlagen über – aus archivischer Sicht – kurze Zeiträume ist teuer genug, wie eine 1996 angestellte Schätzung des NARA Center for Electronic Records für die Archivierung digitaler Unterlagen über 10 Jahre zeigt:<sup>2</sup>

Kosten in US-\$	Kostenart
600.000	6 Arbeitplatzrechner mit je 2 Magnetband- und -kassettenlaufwerken
928.000	Wartung
30.000	2 Prüfstationen für Magnetbandspulen
92.160	Wartung
50.000	2 Prüfstationen für Magnetbandkassetten
46.600	Wartung
10.786.260	Erste Kopieraktion des Gesamtbestands
29.876	10 Prüfläufe (einer pro Jahr)
1.030.640	Weitere Kopier- und Umformatierungsläufe
13.594.396	Gesamtkosten

Die Schätzung geht von einem Bestand von 23.000 digitalen Einheiten und einem durchschnittlichen jährlichen Zugang von 2900 Einheiten aus, ohne allerdings den Umfang einer Einheit anzugeben. Der Löwenanteil der Kosten entfällt auf Kopieraktionen (Refreshing), Migrationen sind hier noch nicht vorgesehen. Und da der Datenbestand ständig wächst, wird auch der Umfang der Kopieraktionen ständig größer, denn diese umfassen ja stets den Gesamtbestand.

Betrachtet man nun Migrationen, so zeigen sich weitere kostenträchtige Faktoren. So muss jeder Migrationslauf sorgfältig protokolliert werden, damit die Authentizität der Unterlagen

<sup>&</sup>lt;sup>1</sup> Charles M. Dollar, Authentic electronic records. Strategies for long-term access, Chicago 2000. – Vgl. auch Bryan Bergeron, Dark ages II. When the digital data die, Upper Saddle River 2002.

<sup>&</sup>lt;sup>2</sup> Charles M. Dollar (wie Anm. 1) S. 211.

nicht gefährdet wird. Hierbei entsteht eine u.U. durchaus beträchtliche Datenmenge.<sup>3</sup> Auch sie wächst mit jeder weiteren Migration, da die Protokolle früherer Migrationen ebenfalls migriert werden müssen. Neben den eigentlichen Daten entsteht also eine sekundäre Datenlast, die jedoch genau so sorgfältig zu bewahren ist.

Hinzu kommt ein weiteres Problem, das sich erst bei Migrationen zu offenbaren pflegt. Wechselt man z.B. ein Datenbankmanagementsystem, sei es anlässlich eines Updates oder eines Umstiegs auf ein anderes Produkt, so werden Inkonsistenzen und Eingabefehler häufig erst zu diesem Zeitpunkt erkannt. Manuelle Nachbearbeitungen sind die Folge und damit erhebliche Personalkosten. Dem Autor sind zwei Beispiele derartiger Migrationen aus dem Bereich der Wirtschaft bekannt; in beiden Fällen überstiegen die tatsächlichen Kosten den Voranschlag um mehr als das Zehnfache.

Die Aufbewahrung digitalen Archivguts in digitaler Form wird also mit steigenden Kosten verbunden sein. Einsparungen, die sich aus dem im Vergleich zu Schriftgut auf Papier geringeren Platzbedarf ergeben, werden diese Kosten nicht wett machen. Man darf nicht übersehen, dass auch die Lagerung von Datenträgern Anforderungen an das Raumklima stellt. Dazu kommt die Notwendigkeit, mindestens eine Sicherungskopie des Datenbestands vorzuhalten. Bei Migrationen bedeutet dies, dass es zweier Läufe bedarf: Zuerst wird die "Masterkopie" migriert, und erst nach der Überprüfung des migrierten Bestands auf Vollständigkeit und Korrektheit kann die Sicherungskopie migriert werden.

Auch die Benützung digitalen Archivguts ist mit einigem Aufwand verbunden. Die Vorstellung, dass ein Ausheber in das Datenmagazin geht, den oder die Datenträger heraussucht, in ein Laufwerk einlegt und nach dem Auslesen der Daten wieder einstellt, ist aus organisatorischer und konservatorischer Sicht mehr als bedenklich. Ein falsch eingestellter Datenträger dürfte so unauffindbar sein wie ein verreihter Aktenband, und schon heute haucht so mancher Datenträger sein Leben aus, wenn er mit Hast und wenig Gefühl ins Laufwerk gedrückt wird. Auf der sicheren Seite ist man mit Jukeboxen oder gar Robotern, doch diese sind wartungsträchtig und von begrenzter Lebensdauer. Es ist wohl nicht zu gewagt, wenn man unter diesen Gesichtspunkten die Kosten für eine Aufbewahrung in digitaler Form höher veranschlagt als für die Lagerung und Bereitstellung klassischen Archivguts. Dann erscheint es jedoch fraglich, ob die für eine digitale Archivierungsform benötigten und kontinuierlich ansteigenden finanziellen Ressourcen von den Archivträgern überhaupt bereit gestellt werden können und ob man willens ist, dies zu tun.

Gibt es Möglichkeiten, digitale Unterlagen in einer anderen Form aufzubewahren, die dem Archiv diese – aus seiner Sicht – kurzfristigen Migrationsschübe erspart? Charles Dollar nennt zwar die Überlieferung auf Papier oder Mikrofilm als Surrogat, wenn es denn gar nicht anders zu machen sei, doch müsse in diesem Fall die Authentizität der digitalen Vorlage eigens neu geschaffen werden, und viele Funktionen könnten auf das Surrogat nicht abgebildet werden. Beispielhaft für derartige Funktionen ist der Hyperlink. Wenn er auf Film (das Surrogat Papier wird man bis auf wenige Ausnahmen nicht in Betracht ziehen) abgebildet ist, kann man ihn nicht mehr anklicken. Doch wie wird es sein, wenn man das digitale Pendant nach einigen Jahrzehnten am Bildschirm betrachtet und den Link aktiviert? Wohin wird er führen? Vermutlich ins Nichts.

Ein anderes Beispiel ist eine Datenbank. Man kann sie auf ein plattformunabhängiges flat file umsetzen, das recherchierbar bleibt, sofern man die Strukturinformationen, Kodierungen und Systemdokumentationen des Datenbanksystems dazu aufbewahrt. Theoretisch könnte man mit Hilfe dieser Informationen das flat file in ein aktuelles Datenbanksystem konvertieren,

<sup>&</sup>lt;sup>3</sup> Das von Charles M. Dollar (wie Anm. 1) S. 231–248 vorgestellte "Preservation metadata model" veranschaulicht die Menge dieser Daten.

<sup>&</sup>lt;sup>4</sup> Charles M. Dollar (wie Anm. 1) S. 73 f.

das zumindest Recherchen und andere Auswertungen ermöglicht. Der Aufwand dürfte allerdings beträchtlich sein.<sup>5</sup>

Digitale Audio- und Videodateien auf das Surrogat Mikrofilm abzubilden ist zwar theoretisch möglich, aber wohl kaum sinnvoll. Doch zumindest im künftigen digitalen Verwaltungsschriftgut dürften solche keine nennenswerte Rolle spielen; hier dominieren wohl auch weiterhin Texte, Grafiken und Fotos. Das Bundesarchiv hat bekanntlich empfohlen – und diese Empfehlungen sind in das DOMEA-Konzept des Bundes eingegangen – Primärinformationen vor der Übergabe an das Archiv in das TIF-Format zu konvertieren und die Metadaten in Nur-Text-Dateien mit ASCII-Kodierung zu übergeben. Doch was ist eine TIFF-Datei anderes als ein digitales Foto? Der dort abgebildete Text liegt nicht mehr als Zeichenfolge, sondern als Bitmap vor. Damit enthält ein TIFF-Bild nicht mehr Informationen als eine von der Auflösung her adäquate Mikrofilmaufnahme. Was spricht also dagegen, auch die Metadaten auf Film auszubelichten, und zwar in einer Qualität, die eine fehlerfreie Rückdigitalisierung gewährleistet? Denn nur in den Metadaten kann recherchiert werden, nicht aber im TIFF-Bild, das eigentlich auch nichts anderes als ein Surrogat der ursprünglichen Word- oder Excel-Datei ist.

Die Technik für das automatisierte Ausbelichten digitaler Daten auf Mikrofilm ist vorhanden. So bietet z.B. Kodak den "Document Archive Writer" an, und die Firma Microbox hat ein Gerät namens Polycom entwickelt, das ein konfigurierbares Selektieren und Herausziehen von Primärinformationen und Metadaten und eine strukturierte Abbildung auf Film erlaubt.

Die Benützung mikroverfilmter digitaler Unterlagen ist aufwändiger und umständlicher als der Umgang mit der digitalen Form. Doch der Mikrofilm bedarf keiner Migration. In größeren zeitlichen Abständen wird es notwendig sein, den Film zu kopieren oder – um die dabei unvermeidlichen Qualitätsverluste zu vermeiden – ihn zu redigitalisieren und davon einen neuen Film zu ziehen. Die Kosten hierfür werden nicht gering sein, doch sie sind kalkulierbar und fallen vielleicht alle 100 Jahre an. Laufende Kosten fallen natürlich für die sachgerechte Lagerung der Filme an, und so fragt man sich, ob es nicht andere Möglichkeiten gibt, ein gegenüber Umwelteinflüssen weitgehend unempfindliches und jederzeit mit einfachen Techniken lesbares Speichermedium zu erzeugen, das große Datenvolumina aufnimmt und dadurch wenig Platz beansprucht.

Schon vor etlichen Jahren gab es Pressemeldungen, wonach in Kalifornien eine Art CD aus feuer- und säurefestem Material (Iridium) entwickelt werde, in die man digitale Daten und Informationen zu ihrer Dekodierung in extremer Verkleinerung "eingravieren" könne. Die digitalen Daten sollten in binärer Kodierung gleichsam "versteinert" werden, die Dekodierungsinformationen hingegen als analoger Klartext. Diese "CD" sollte etwa 650 Gigabyte Daten aufnehmen können. Weiteres ließ sich zu diesem Projekt nicht in Erfahrung bringen, die Idee wurde jedoch anderweitig aufgegriffen.

Die vor einigen Jahren in den USA gegründete LongNow Foundation<sup>7</sup> hat sich zum Ziel gesetzt, wichtige Schriftzeugnisse menschlicher Kultur über 10.000 Jahre hinweg zu bewahren (10kLibrary). Hinzu kam der Wunsch, diese Zeugnisse auch in sprachlicher Vielfalt zu sichern. Man wählte rund 1000 Sprachen aus und entwickelte für jede ein Informationspaket, das künftigen Linguisten die Möglichkeit geben soll, die Sprache auch nach ihrem Untergang zu rekonstruieren. Als Textbeispiel, das in jeder dieser Sprachen vorliegt, dienen die ersten drei Kapitel der Genesis.

Für einen ersten Versuch entschied man sich für eine Dokumentation der 1000 Sprachen, die jeweils 27 Druckseiten umfasst, also insgesamt 27.000 Seiten. Hinzu kam der vollständige

\_

<sup>&</sup>lt;sup>5</sup> Vgl. den Beitrag von Björn Dehms, Andreas Engel und Ulrich Meyer in diesem Band.

<sup>&</sup>lt;sup>6</sup> Dies ist z.B. durch die Wahl eines OCR-Fonts (OCR-A oder OCR-B) vor dem Ausbelichten zu erreichen.

<sup>&</sup>lt;sup>7</sup> Vgl. http://www.longnow.org/10klibrary/rosetta.html

Text der Genesis in acht wichtigen Sprachen mit etwa 3000 Seiten. All dieses sollte auf einem kleinen Speichermedium Platz finden, und zwar in analoger und beständiger Form, zu lesen mit einem optischen Mikroskop mit dem Vergrößerungsfaktor tausend. Das Medium bekam den Namen HD-Rosetta: HD steht für hard disk und spielt auf die Robustheit des Mediums an, Rosetta auf die parallele Textüberlieferung.

Tatsächlich ist es der Firma Norsam Technologies<sup>8</sup> gelungen, einen diesen Anforderungen genügenden Datenträger zu entwickeln. In der jetzigen Ausprägung handelt es sich um eine quadratische Nickelplatte mit 5 cm Kantenlänge und einer Dicke von knapp über 6 mm. Daten werden in analoger Form in einer Seitengröße von 21,59 mal 17,9 cm (bei normaler Vergrößerung) abgebildet. Wenn der Datenträger mit optischen Mikroskopen gelesen werden soll, fasst er 5000 bis 18.000 Seiten, steht jedoch ein Elektronenmikroskop zur Verfügung, so können etwa 196.000 Seiten auf dem Medium untergebracht werden. Der Prototyp eines Lesegeräts, das das Mikroskopbild auf einen Computermonitor überträgt (also redigitalisiert!), sollte zum Verkaufspreis von etwa 10.000 US-\$ in Serie gehen. Da die HD-Rosetta jedoch ursprünglich rund war und die Abbildungen in Form einer Spirale angeordnet waren, was sich nicht bewährt hat, wechselte man zur quadratischen Form, die auch eine einfache Navigation über ein Koordinatensystem ermöglicht. Hierfür wird zur Zeit ein spezielles Lesegerät entwickelt.

Unempfindlichkeit und Haltbarkeit des Mediums sind beachtlich. Das Material ist im Wesentlichen Nickel, das gegen elektromagnetische Felder unempfindlich ist (also auch gegen die berüchtigte "elektromagnetische Bombe"). Testberichte des Los Alamos National Laboratory zeigen, dass Temperaturen bis zu 300° C keine Veränderungen des Mediums hervorrufen; erst ab 450° C oxidierte die Oberfläche und machte die Daten teilweise unlesbar. Als man die HD 15 Wochen lang Salzwasser, Leitungswasser oder Meeresluft aussetzte, war keine die Lesbarkeit beeinflussende Korrosion festzustellen. Die Lebensdauer des Mediums wird für normale klimatische Bedingungen mit mehreren tausend Jahren angegeben.

Die HD-Rosetta wird mit einem fokussierten Ionenstrahl beschrieben, der Galliumionen auf die Nickeloberfläche schießt. Die Ionen ätzen Atome aus der Oberfläche heraus und "gravieren" damit Vertiefungen. Als Datenquelle dienen Mikrofilme oder TIFF-Dateien. Das Schreiben einer Seite dauert weniger als eine Zehntelsekunde; in zwei Stunden sind etwa 7000 Seiten eingeätzt. Den Schreibprozess kann derzeit nur die Firma Norsam durchführen; das Verfahren ist patentiert.

Leider gibt die Firma auf ihrer Webseite keine Hinweise auf Praxistests mit Anwendern und nennt auch keine Referenzen. Es bleibt also abzuwarten, ob und wann das Verfahren auf den Markt gelangen wird.

Der Traum vom Archiv, das in einen Stülpdeckelkarton passt und bedenkenlos überall hingestellt werden kann, bleibt also noch unerfüllt. Man sollte diese und andere Entwicklungen jedoch sorgfältig und unvoreingenommen beobachten.

\_

<sup>&</sup>lt;sup>8</sup> Vgl. *http://www.norsam.com* . Leider erhielt der Autor trotz mehrerer Anfragen – schriftlich und als E-Mail – bis heute von der Firma keine Antwort.