

KOST.07

Kolloquium „Archivtaugliche Speicherinfrastruktur“

Thema: Datenauslagerung

Inhalt

1	Einleitung	2
2	Referate	2
2.1	Peter Fornaro, imaging & media lab der Universität Basel	2
2.2	René Meier, Marketing & Sales, Swiss Data Safe AG	3
3	Diskussion	4
3.1	Wie werden die archivischen Anforderungen erfüllt?	4
3.2	Mit welcher Grössenordnung ist für die Kosten zu rechnen?.....	4
3.3	Realisierbarkeit: Ist das etwas für die Staatsarchive?.....	4

1 Einleitung

Die vierte Veranstaltung des KOST-Kolloquiums "Archivtaugliche Speicherinfrastruktur" befasste sich mit dem Thema "Datenauslagerung". Dabei werden digitale Daten auf ein langzeitstabiles Speichermedium umkopiert und in dieser Form aus der operativen Informatikinfrastruktur ausgelagert. Während der Phase der Auslagerung fallen nur die Kosten für die Lagerung des Trägermediums an. Diese Lösung wird schon länger diskutiert und eingesetzt (im COLD Archiving [Computer Output To Laser Disk], vgl. aber auch den inzwischen obsoleten *Photodigital Store* von IBM¹, 1967), erlangt aber mit dem Einsatz von Mikrofilm als Trägermedium eine neue Dimension bezüglich Auslagerungszeitraum (> 100 Jahre).

Das Thema wurde im KOST-Kolloquium aus zwei verschiedenen Blickwinkeln beleuchtet: Herr Peter Fornaro vom imaging & media lab der Universität Basel² stellte zunächst an Hand des Nationalfonds-Projekts PEVIAR (Permanent Visual Archive)³ den aktuellen Forschungsstand; Herr René Meier von der Firma Swiss Data Safe⁴ präsentierte anschliessend mögliche kommerzielle Anwendungen dieser Forschungen.

2 Referate

2.1 Peter Fornaro, imaging & media lab der Universität Basel

Herr Fornaro gab zunächst einen kurzen Überblick über grundlegende Probleme der digitalen Archivierung. Er beleuchtete zwei Ansätze, nämlich die (regelmässige) Datenmigration und die Verwendung von langlebigen Speichersystemen. Letztere empfiehlt sich vor allem für Daten mit geringer Lesewahrscheinlichkeit.

Daran knüpft PEVIAR an. Das Konzept verbindet einen stabilen Träger, möglichst langlebige Dateiformate und eine einfache Zugriffsmethode. Mit Laserbelichtungsgeräten werden analoge und digitale Inhalte auf Farbmikrofilm gebracht. Das Lesen kann im Prinzip mit jedem verfügbaren Scanner erfolgen. Deshalb erübrigt sich das Aufbewahren einer geeigneten Lesehardware; oder anders gesagt: Der Auslagerungszeitraum wird weder durch die Verfügbarkeit eines bestimmten Lesegerätes (im Gegensatz beispielsweise zu magneto-optischen Trägern oder DVDs); noch durch die kurze Lebensdauer des Trägermediums beschränkt.

PEVIAR arbeitet mit einem hybridfähigen Speichermedium: Auf dem Mikrofilm können sowohl analoge Inhalte in menschenlesbarer Form (Bilder, Metadaten im Textformat) als auch digital codierte Informationen (in einem zweidimensionalen Barcode) gespeichert werden. Analoge Inhalte sind bereits mit rudimentären Hilfsmitteln (Lupe) vom Menschen les- und verstehbar. Digitale Daten müssen erst durch Redigitalisieren des gespeicherten Bitmusters mit einem Decoderprogramm in eine Datei zurückverwandelt werden (z.B. TIFF); anschliessend kann die Datei mit einem entsprechenden Viewer betrachtet werden. Hier begrenzen also die Verfügbarkeit von Decoder und Viewer die

¹ http://en.wikipedia.org/wiki/IBM_1360

² <http://www.abmt.unibas.ch/>

³ <http://www.peviar.ch/>

⁴ <http://www.swissdatasafe.ch/>

Auslagerungszeit. Allerdings wäre es im Prinzip denkbar, die Decoder- und Vieweranleitung in Klartext auf dem Mikrofilm zu speichern und erst im Anwendungsfall neu zu programmieren.

Der Vorteil der digitalen Speicherung parallel zur analogen liegt darin, dass Metadaten, Layer und Perspektiven z.B. bei GIS-Daten im digitalen Bild erhalten bleiben. Redundanz in der Codierung (*error correction mechanisms*) schützen vor Kratzern und Beschädigungen.

Eine Standardisierung der digitalen Codierung und des so genannten Informationslayouts auf dem Film (Anordnung von digitalen und analogen Frames) ist noch nicht erreicht; es gibt erst Versuche in dieser Richtung.

Die Punktgrösse der digitalen Codierung wird durch die Auflösung des Lesegeräts bestimmt: Diese muss mindestens um einen Faktor 3 grösser sein als diejenige des Schreibgeräts. Heutige Highend-Lesegeräte können Punktgrössen von 1-2 μ (Mikrometer) auflösen, was für den Schreibvorgang eine Punktgrösse von 5-6 μ ergibt. Damit können theoretisch auf einen 35-mm-Film von 600 m Länge (entspricht etwa einer Rolle von 35 cm Durchmesser) ungefähr 250 GB Nutzdaten verlustfrei komprimiert geschrieben werden. Zu beachten ist, dass die Speicherdichte im Quadrat zur Verkleinerung der Punktgrösse steigt. PEVIAR verwendet zurzeit Farbmikrofilm mit drei Farblayern; durch weitere Layer könnte die Speicherdichte weiter erhöht werden. Das IML arbeitet mit dem Mathematischen Institut der Universität Zürich⁵ auch an einer Kodierung mit Hilfe der Farbsättigung der Pixel. Zudem entwickelt sich auch die Imaging-Technologie ständig weiter, was sich für den Laien beispielsweise an der Pixeldichte von Digitalkameras ablesen lässt. Denkbar wäre schliesslich, dünneres Filmmaterial, z.B. in einer Kassette einzusetzen.⁶

2.2 René Meier, Marketing & Sales, Swiss Data Safe AG

Swiss Data Safe bietet auf der Basis der Methode "*Bits on Film*" hauptsächlich zwei Dienstleistungen an. Einerseits stellt die Firma Einmal-Ausbelichtungen her als Sicherheitskopien in festgesetzten Abständen, die danach beim Kunden aufbewahrt und im Notfall wieder eingelesen werden. Andererseits bietet sie auch ein vollständiges Langzeitarchiv an, wobei sie für den Kunden auch die ISO 17799 (Information technology -- Code of practice for information security management)⁷ übernimmt. Dabei bietet "*Bits on Film*" eine Sicherheitskopie, nicht aber die Gebrauchskopie der Daten. Der Datenfilm ist ausdrücklich als ein Offline-Speichermedium gedacht, auf dem Daten archiviert werden, so das Online-System entlasten und zu einem späteren Zeitpunkt wieder in ein zukünftiges Online-System eingebracht werden können. Die Sicherungskopie sollte mindestens zehn Mal länger als die Gebrauchskopie halten. Belichtet wird nach einem proprietären Informationslayout auf unperforierten 35-mm-Film.

Die Dienstleistung für einen zertifizierten Langzeitarchivierungsservice ist sehr kundenspezifisch und daher individuell zu berechnen; ein generisches Kostenmodell besteht

⁵ <http://www.math.unizh.ch/aa/index.php?id=56>

⁶ Es muss allerdings davon ausgegangen werden, dass ein Äquivalent zum Mooreschen Gesetz (http://de.wikipedia.org/wiki/Mooresches_Gesetz) für Mikrofilm nicht existiert, dass die Speicherdichte also nicht so schnell zunimmt wie die Datenmengen.

⁷ http://www.iso.org/iso/catalogue_detail?csnumber=33441; siehe dazu auch http://de.wikipedia.org/wiki/ISO_17799.

nicht. Als Richtwert kann man aber von einmaligen Kosten für die Ausbelichtung von 1 TB Daten von rund 25'000 EUR ausgehen.

3 Diskussion

3.1 Wie werden die archivischen Anforderungen erfüllt?

Gewisse archivische Anforderungen werden von *"Bits on Film"* vorzüglich erfüllt. Das Trägermedium ist äusserst dauerhaft (500 Jahre⁸); weit dauerhafter, als in der Regel notwendig. Es besteht keine direkte Abhängigkeit von Hard- und Software, wenn wir Scannen oder digitales Fotografieren weiterhin als eine Basistechnologie voraussetzen. Die Migrationsrate und damit die Migrationskosten sind gering, wenn durch die Wahl geeigneter Dateiformate eine Formatmigration frühestens in 30 Jahre notwendig wird. Problematischer ist die Überprüfung der Integrität des archivierten Bitstroms: Nach dem Belichten werden die Informationen zur Kontrolle nochmals eingelesen; danach ist jedoch noch keine einfache, automatisierbare Integritätsprüfung möglich.

Durch das Einfügen von analogen Informationen im Informationslayout können wir im Falle einer sehr langen Lagerung eine Überlieferung auch über unsere heutige Informatik hinaus gewährleisten.⁹

3.2 Mit welcher Grössenordnung ist für die Kosten zu rechnen?

Swiss Data Safe AG schreibt ein TB für 25'000 Euro in einem eigenen Format auf Film. Für diesen Betrag erhält der Kunde den belichteten und durch anschliessendes Einlesen verifizierten Film. Ein allfälliges Rückdigitalisieren zu einem späteren Zeitpunkt kann der Kunde sowohl selber übernehmen als auch von Swiss Data Safe ausführen lassen. Alle Konstruktionsdaten (Dekodierungssoftware, Informationslayout auf dem Film, Beschreibung des Lesegeräts) sind notariell hinterlegt und können dem Kunden so im Bedarfsfall zugänglich gemacht werden.

Herr Fornaro macht darauf aufmerksam, dass bei den heutigen Kostenverhältnissen ein Belichten von Schwarzweissfilm, auch bei wesentlich geringer Datendichte, die mit Abstand günstigste Lösung wäre.

3.3 Realisierbarkeit: Ist das etwas für die Staatsarchive?

Diese Technologie steht nicht in Konkurrenz zu klassischen Speichersystemen; entsprechend kommt ihr Einsatz für ein Staatsarchiv wahrscheinlich nur ergänzend zu anderen Lösungen in Betracht, da der Zugriff noch in keiner Weise automatisiert und damit sehr aufwendig ist, da die Leserate gering ist und lange Zugriffszeiten in Kauf genommen werden müssen. Als ultimative Backuplösung für eine bestehende Online-Speicherung bietet *"Bits on Film"* aber interessante Aspekte. Zurzeit hat die Technologie ihre Maturität noch nicht erreicht. Eine Standardisierung und Normierung wäre für die Archive von grösster Bedeutung und würde die Akzeptanz der Methode wesentlich erhöhen.

⁸ "Die Haltbarkeit von Mikrofilm soll bei entsprechender Lagerung (21 °C, 50 % relative Luftfeuchte) bis zu 500 Jahre betragen." (Quelle: <http://de.wikipedia.org/wiki/Mikrofilm>)

⁹ Siehe Diskussionen um die Rosetta Disk, <http://de.wikipedia.org/wiki/Rosetta-Projekt>.