

**Eine ARCHE zur Rettung digitalen Archivguts -
Langzeitarchivierung digitaler Daten auf laserbelichtetem
Farbmikrofilm**

Universitätsbibliothek Stuttgart / Landesarchiv Baden-Württemberg

Klaus Wendel

Klaus.Wendel@la-bw.de

Vortrag gehalten auf dem 95. Deutschen Bibliothekartag in Dresden am 24.3.2006

1 Grundprobleme der Langzeitarchivierung digitaler Daten

Wir finden Großvaters Schuhkarton auf dem Dachboden. Briefe, Zeugnisse, Urkunden und eine Vielzahl Fotos. Überwiegend schwarz/weiß. Sogar einige Farbfotos sind darunter. Ein leichter Blaustich läßt ihr Alter vermuten. Auch das Papier zeigt deutliche Alterungserscheinungen. Doch trotz des Zerfalls sind die Informationen noch fast ausnahmslos les- oder zumindest erschließbar. Versetzen wir den Fund nun um ein paar Jahre in die Zukunft. Ratlos stehen unsere Enkel vor unzähligen Disketten im 5^{1/4}" und 3^{1/2}"-Format, verschiedenartigen Ton- und Datenbändern und vor allem CD-ROMs. Das heißt diversen Silberscheiben mit 650, 4500 oder 9000 MB Kapazität. Ganz zu schweigen von den unzähligen anderen technischen Innovationen, welche es bis dahin mit Sicherheit gegeben haben wird. Den (übrigens sehr unwahrscheinlichen) Fall vorausgesetzt, daß all diese Datenträger ihre physikalischen Eigenschaften nicht in der Zwischenzeit verloren haben, so erforderte es jetzt ein ganzes Technikmuseum, das die als Lesegeräte notwendigen Computer, Diskettenlaufwerke, Streamer und CD-Reader einschließlich der benötigten Betriebssoftware in funktionsfähigem Zustand bereit hielt. Und man wüßte im Voraus noch nicht einmal, welche Daten die undurchsichtigen Medien überhaupt beinhaltenen und ob sie den hohen Aufwand einer Rekonstruktion überhaupt rechtfertigten. In jedem Fall müßten sich die Nachkommen damit abfinden, daß ein Teil der Daten für immer verloren wäre und daß der Rest nicht unmittelbar, sondern nur mit erheblichem Kostenaufwand zurückgelesen werden könnte.

Wir sehen uns bei der Langzeitspeicherung digitaler Daten also folgenden Grundproblemen konfrontiert:

1. Je flüchtiger ein Medium, desto höher die Gefahr eines Migrationsbruchs. (Bei digitalen Daten bedeutete dies den Totalverlust der Information!)
2. Datenmigration gestaltet sich schon nach kurzer Zeit zunehmend kostspielig.
3. Möglichkeiten der Datenmigration sind schon für die nahe Zukunft unüberschaubar. Der Erhalt der Daten ist somit nicht kalkulierbar.

Kurioser Weise würden zukünftige Generationen aus dieser Sammlung selbst dann noch einen größeren Nutzen ziehen können, wenn die Informationen auf Pergament gemalt oder in Marmor gemeißelt worden wären. Mit anderen Worten: Der nur mit speziellen Werkzeugen *lesbare* digitale Master hat seine Bedeutung eingebüßt, wohingegen der nur mit geringem technischen Aufwand *interpretierbaren* Information ein steigender Wert zukommt. Die Lösung des Problems liegt also in der Wahl eines Mediums mit erheblich längeren Migrationsintervallen.

Nun eignen sich Tierhäute und Stein nicht sehr gut zur Langzeitarchivierung heutiger Datenmengen. Ideal wäre vielmehr ein Medium, das über eine hohe Auflösung verfügt, seine Farbechtheit auch über einen langen Zeitraum nicht einbüßt, maschinell beschreibbar ist, beliebige Datenformate aufnimmt und zudem nicht ganz soviel Raum beansprucht, wie beispielsweise besagte Marmorplatten. Wir verfügen über dieses Medium bereits: Den Mikrofilm.

Bedeutungsgegensatz

Originalität des Mediums - Information

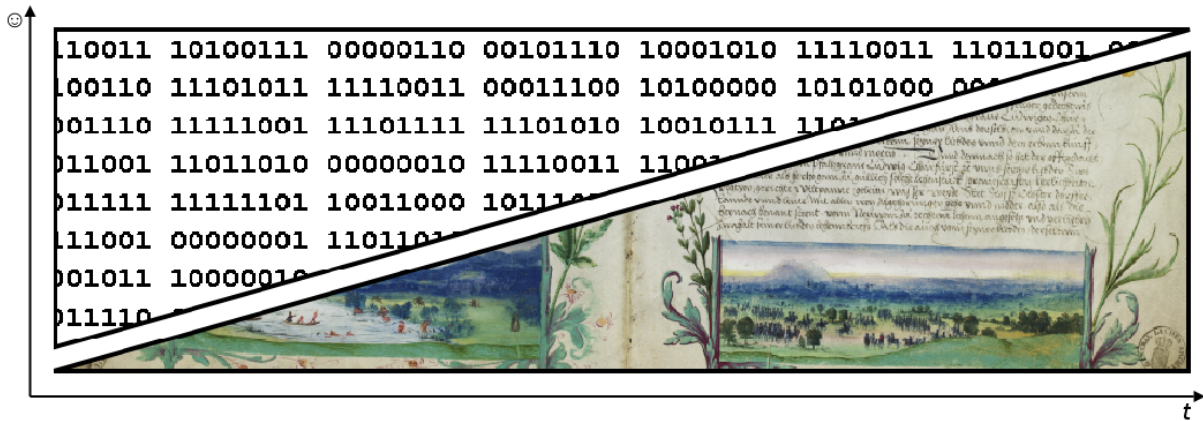
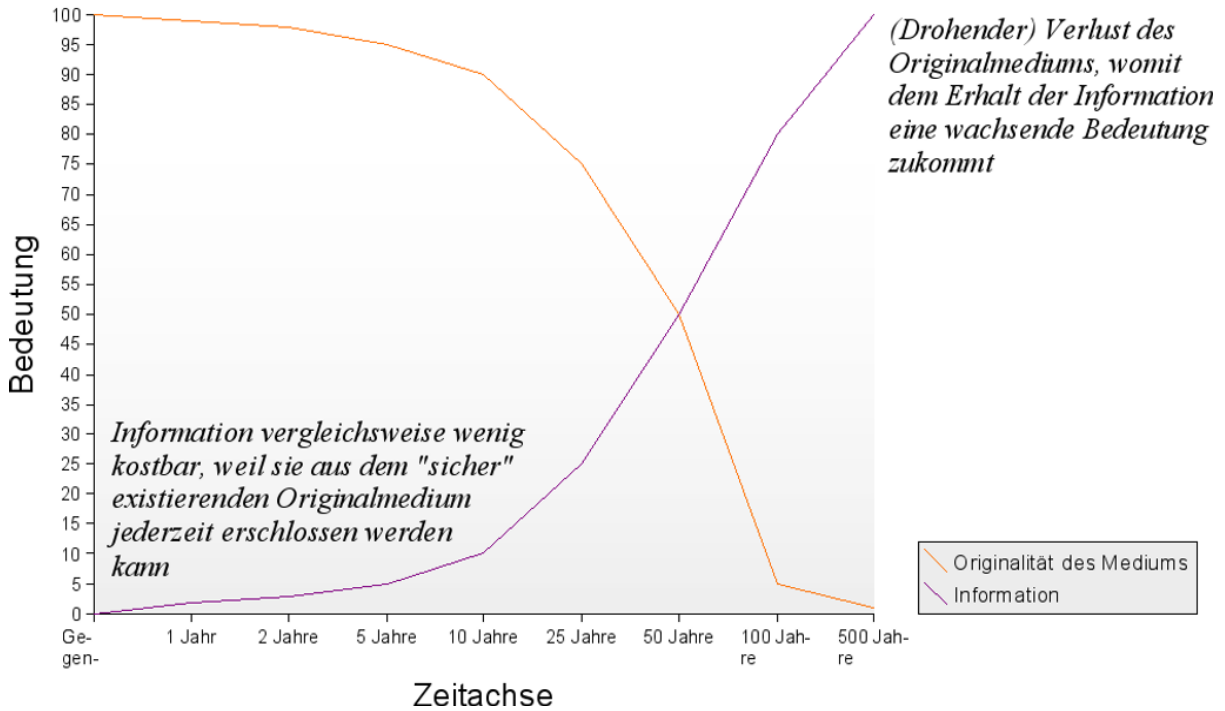


Abbildung 1: Bedeutungsgegensatz: Originalität des Mediums \Leftrightarrow Information



Abbildung 2: Das ARCHE-Projekt und seine wichtigsten Partner

2 Das Verbundprojekt ARCHE

Das durch das Bundesministerium für Wirtschaft und Arbeit im Rahmen des InnoNet¹ geförderte Verbundprojekt ARCHE nutzt diesen klassischen Mikrofilm. Neu sind allerdings das Verfahren mit welchem er belichtet wird und die qualitativen Verbesserungen, die sich dadurch ergeben. Im Mittelpunkt des Projektes steht der Farbmikrofilm-Laserbelichter, den das Fraunhofer-Institut für Physikalische Meßtechnik (IPM) in Freiburg in Zusammenarbeit mit der Frankfurter Firma Microarchive Systems (MAS) entwickelt hat.

3 Die Technik

Neben der herkömmlichen *analogen* Fotografie, die ein durch ein Objektiv verkleinertes Bild auf den Film belichtet, existieren bereits Verfahren, welche digitale Bilder über den Umweg einer Projektion analog verfilmen. Demgegenüber hat der in Freiburg entwickelte Laserbelichter den ganz erheblichen Vorzug, durch Nutzung von Lasern mit höchster Strahlqualität keine Abbildungsfehler akzeptieren zu müssen, die sich sonst selbst bei den besten Optiken einstellen. Die direkte Ansteuerung der Laser ermöglicht ein sehr konstantes und höchst genaues Farbmanagement mit 36 Bit Farbtiefe. Insgesamt drei den Farbkanälen rot, grün und blau entsprechend getaktete einfarbige Laserimpulse werden hier zu einem bunten Lichtstrahl gebündelt und als Bildpunkt direkt auf den Film geschrieben. Damit wurde eine Pixelgröße von 3 μm möglich, was einer Auflösung von 160 Linienpaaren / mm entspricht. Gemessen an den 32x45 mm großen „Frames“ eines 35 mm Films erreicht man eine Auflösung von 10666 x 15000 Bildpunkten auf äußerst beständigem Ilfochrome[®] (Cibachrome) Mikrofilm.

4 Die Software

Doch von vornherein sollte ARCHE mehr leisten, als nur digitale Bilder in möglichst hoher Qualität auf den Film zu kopieren. Auch die Metadaten, also die zugehörigen Findbücher und Datenbankauszüge, sollten mit auf den Film. Und zwar in einer Form, die später einmal eine möglichst unverfälschte und automatisierte Redigitalisierung gestattet. Dazu haben das Landesarchiv Baden-Württemberg und die Universitätsbibliothek Stuttgart den Workflow definiert und die hierfür erforderliche Software entwickelt. Eine wichtige Herausforderung stellte dabei die große Anzahl unterschiedlicher Datenbanksysteme dar. Schließlich wurde eine universelle Schnittstelle gefunden, die jede beliebige Datenkombination unter Beibehaltung ihrer redaktionellen Struktur aufnehmen kann. Die Speicherung der Metadaten auf dem Film erfolgt nun sowohl in menschenlesbarer Textform, als auch in maschinenlesbarem XML-Format. Überhaupt wurde bei der Aufbringung der Daten größten Wert darauf gelegt, daß die spätere Rückgewinnung nicht zwingend von einer

¹ Förderung von innovativen Netzwerken
(http://www.vdivde-it.de/innonet/projekte/in_pp079_ arche.pdf)

Funktionsweise des ArchivLasers

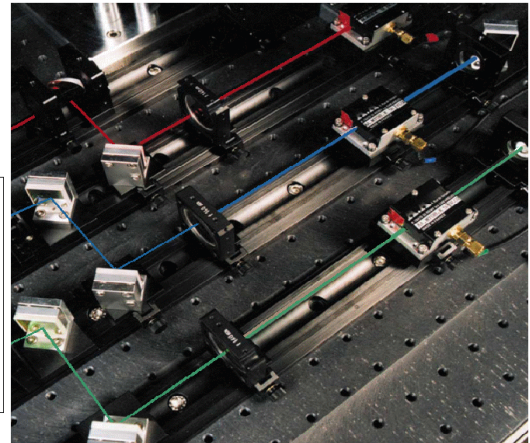
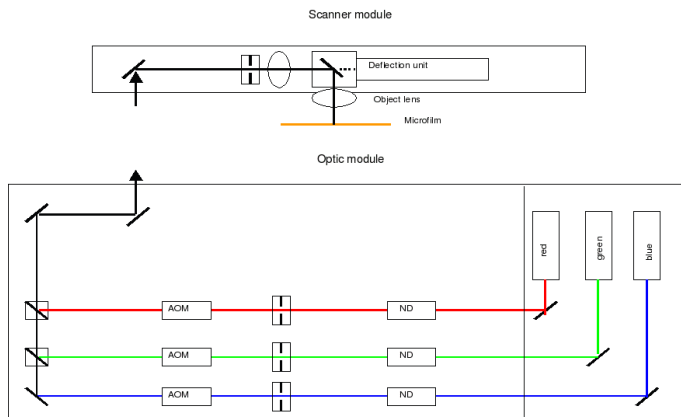


fig. 1

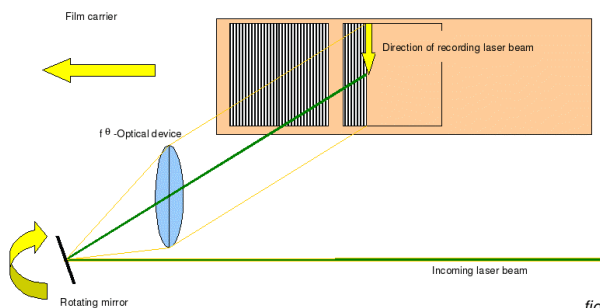
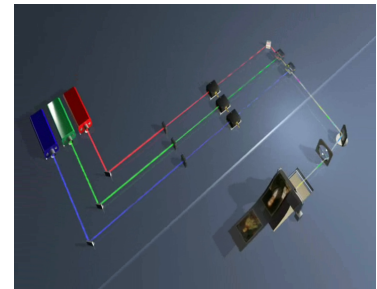
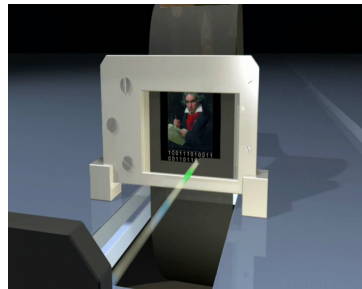
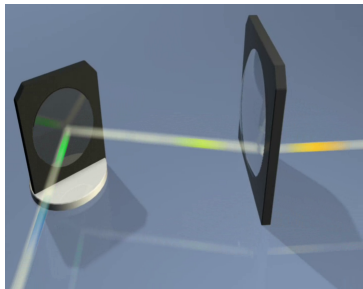
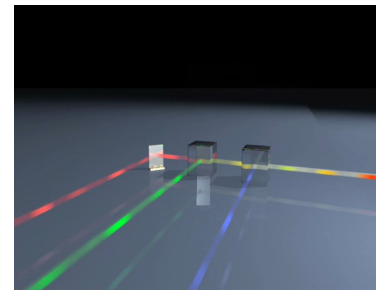
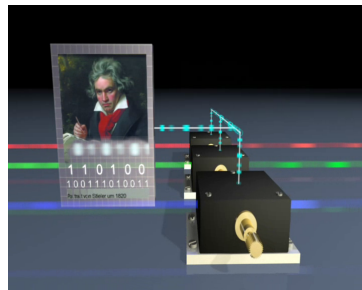
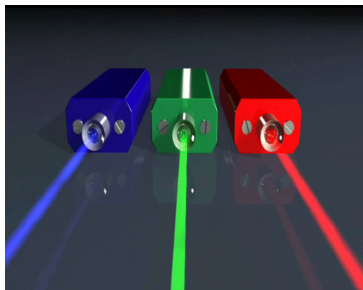


fig. 2



Technische Daten des ArchivLasers

Framegröße	32 x 45 mm (35 mm Film)	Belichtungszeit	40 sec / Frame
Pixelgröße	3 µm (160 Linienpaare / mm)	Filmlänge	≤ 600 m (13,000 Frames)
Pixel / Frame	10,666 x 15,000 Pixel	Zoomfaktor	34 (1 DIN A0 Blatt / Frame)
Farbtiefe	36 (3 x 12) Bit	Kapazität	16 Blätter DIN A4 / Frame bzw. 1 GB / 25 cm Film
Filmmaterial	S/W oder Color-Mikrofilm (Ilford, Kodak, Agfa)	Produktivität	≈ 1 km Film in vier Tagen

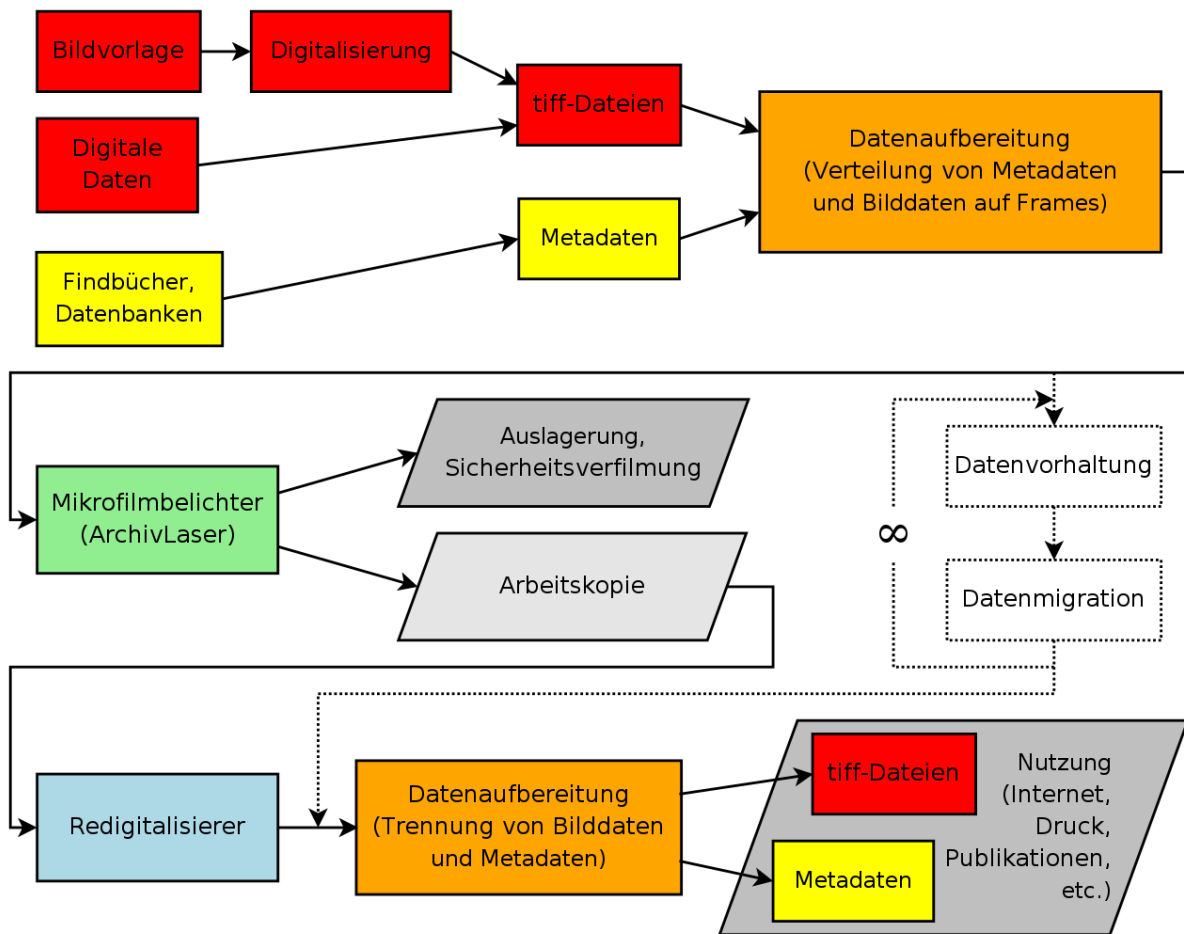


Abbildung 4: Der ARCHE-Workflow von der digitalen Datei bis zur Redigitalisierung

komplizierten Hardware abhängig ist, sondern im Zweifel auch manuell mit Hilfe einfachster optischer Instrumente erfolgen kann. Trotzdem wäre die Ausbelichtung digitaler Strichcodes oder Bitmasken durchaus denkbar.

Die Bilddaten werden in farblich gekennzeichneten Containern gespeichert, die entweder einzeln oder auch kombiniert in einen Frame gestellt werden können. Dieses so genannte „nesting“ kann die Ausnutzung des Mikrofilms ganz erheblich verbessern, denn sehr viele digitale Vorlagen reichen nicht an das theoretische Maximum des ARCHE-Mikrofilms heran. Grenzen setzen hier vor allem die digitalen Speichermedien bei der Aufnahme.



Abbildung 5: Schematische Darstellung eines ARCHE-Mikrofilms

Frame 1: Beschreibung des Bestandes

Frame 2: Datenbankauszug in lesbarer Form

Frame 3: Datenbankauszug in XML-Format

Frame 4-5: Speicherung einzelner Bilder pro Frame in horizontaler und vertikaler Ausrichtung

Frame 6-8: Beispiele für kombinierte Speicherung von Bildern durch „nesting“; Frame 8 enthält Container mit einer mehrfarbigen Bitmaske und einer bitonalen Datamatrix

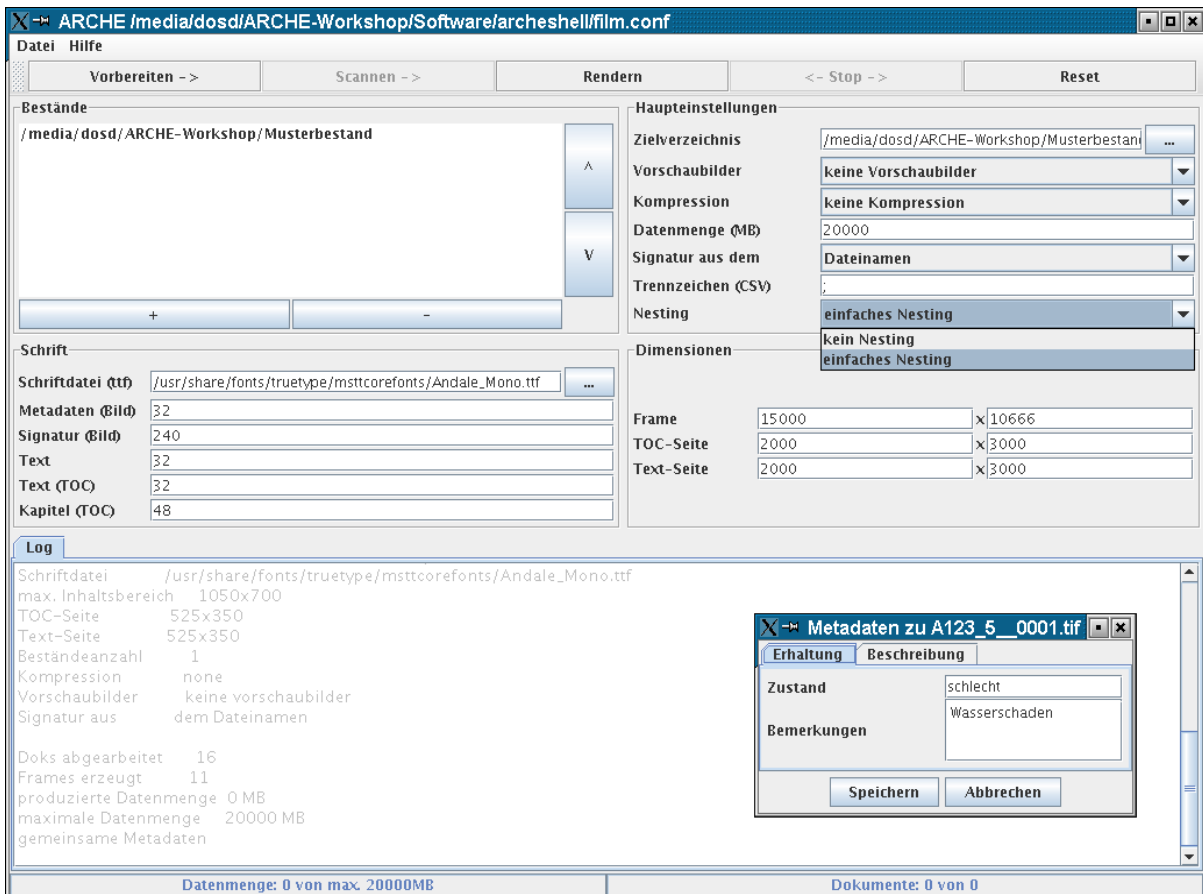


Abbildung 6: „archegui” und „metainput”, zwei Komponenten der in der Universitätsbibliothek Stuttgart von [Wladimir Schwitin](#) entwickelten Softwareschnittstelle zur Bildbearbeitung

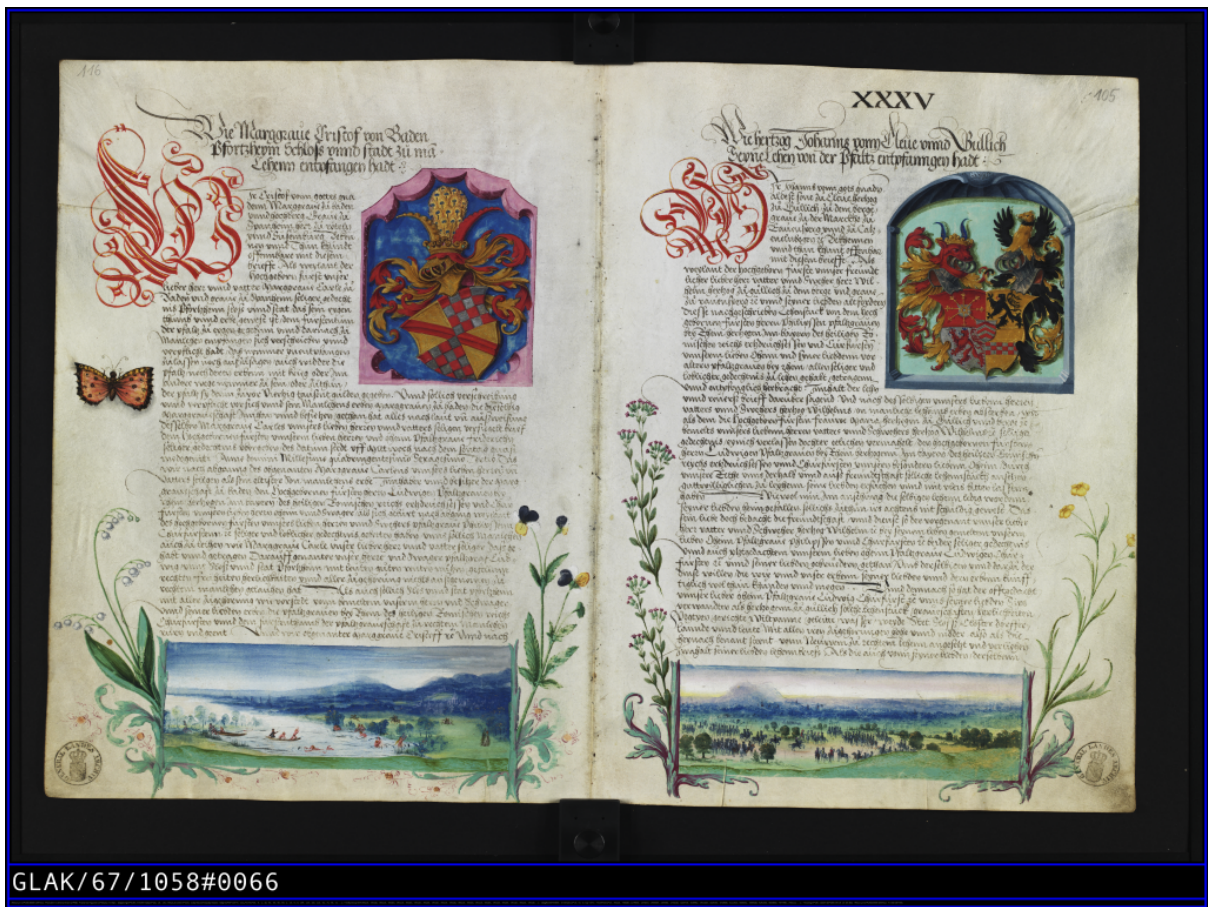


Abbildung 7: Ein Beispiel aus der Praxis: Lehenbuch des Kurfürsten Ludwig V. von der Pfalz (frühes 16. Jahrhundert)

5 Die Praxis

Ein einziges 22 Megapixel (5440 x 4080 Bildpunkte) großes Bild hat bei einer Farbtiefe von 48 Bit ein Volumen von immerhin 127 MB. Die größten im ARCHE-Projekt verarbeiteten Formate waren die mit einem Reproscanner der Firma Homrich erzeugten 87 Megapixel (10880 x 8160 Bildpunkte) großen Aufnahmen des Lehenbuchs des Kurfürsten Ludwig V. von der Pfalz aus dem frühen 16. Jahrhundert. Ein einziges dieser Bilder beansprucht 503 MB Festplattenplatz, der gesamte Band sogar mehr als 300 GB. Solche noch immer kaum zu bewältigenden Datenmengen machen einmal mehr den Nutzen der ARCHE-Verfilmung deutlich.

Das Pfälzische Lehenbuch, welches sich im Besitz des Generallandesarchivs Karlsruhe befindet, war nur ein kleiner Teil eines vielfältig zusammengesetzten Warenkorbes an Archivalien und Bibliotheksgut, mit dessen Hilfe der ARCHE-Workflow nutzbringend durchgespielt und erfolgreich getestet werden konnte. Zu erwähnen sind außerdem Archiv-



Abbildung 8: Signatur und bildbezogene Metadaten in einem Container

bestände des Hauptstaatsarchivs Stuttgart, des Staatsarchivs Ludwigsburg und des Universitätsarchivs Stuttgart. Zu bewältigen waren auch historische Landkarten aus dem Besitz der Universitätsbibliothek Stuttgart und Dias des Stuttgarter Instituts für Leichtbau, Entwerfen und Konstruieren (ILEK). Workflow und Software bewährten sich auch an mehreren in PDF-Format vorliegenden Dissertationen und Habilitationsschriften des Online Publikationsverbundes OPUS der Universität Stuttgart.

6 Ausblick

ARCHE wurde im Sommer 2004 ins Leben gerufen und steht nun kurz vor dem erfolgreichen Abschluß. In dieser Zeit ist es gelungen, den bewährten Mikrofilm einer weiteren Anwendung zuzuführen. Es ist zu erwarten und im Dienste der Bestandserhaltung sehr wünschenswert, daß dieses fast schon totgesagte Medium dadurch eine Renaissance erlebt. Es gibt derzeit nicht viele Alternativen, die in der Lage wären, dem Verlust der heute in großer Zahl gewonnenen Digitaldaten vorzubeugen. Anders als bei herkömmlichem Archivgut muß uns bewußt werden, daß mit fortschreitender Zeit der Erhalt des digitalen Masters seine Bedeutung verliert, dagegen die Bewahrung der kostbaren Information den eigentlichen Wert darstellt!