Digitale Archive und Bibliotheken

Herausgegeben von
Hartmut Weber und Gerald Maier

Kohlhammer

Weber/Maier, Digitale Archive und Bibliotheken

WERKHEFTE
DER STAATLICHEN ARCHIVVERWALTUNG
BADEN-WÜRTTEMBERG

Herausgegeben
von der Landesarchivdirektion
Baden-Württemberg

Serie A Landesarchivdirektion

Heft 15

2000

Verlag W. Kohlhammer Stuttgart

Digitale Archive und Bibliotheken

Neue Zugangsmöglichkeiten und Nutzungsqualitäten

Herausgegeben von
Hartmut Weber und Gerald Maier

2000

Verlag W. Kohlhammer Stuttgart

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Digitale Archive und Bibliotheken : neue Zugangsmöglichkeiten und Nutzungsqualitäten / hrsg. von Hartmut Weber und Gerald Maier. - Stuttgart ; Berlin ; Köln : Kohlhammer, 2000
(Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg :
Serie A, Landesarchivdirektion ; H. 15)
ISBN 3-17-016062-1

Diese Publikation ist auf alterungsbeständigem, säurefreiem Papier gedruckt.

Alle Rechte vorbehalten
© 2000 by Landesarchivdirektion Baden-Württemberg, Stuttgart
Kommissionsverlag: W. Kohlhammer GmbH, Stuttgart
Gesamtherstellung: Verlagsdruckerei Schmidt GmbH, Neustadt an der Aisch
Printed in Germany

Inhalt

Vorwort 9

Hartmut Weber und Gerald Maier
Einleitung 11

Fachkonzepte für digitale Archive und Bibliotheken 19

Sönke Lorenz
Anforderungen der historischen Forschung an einen digitalen Arbeitsplatz 21

Angelika Menne-Haritz
Digitaler Lesesaal, virtuelle Magazine und Online-Findbücher. Auswirkungen der Digitalisierung auf die archivischen Fachaufgaben 25

Carol Mandel
The Transition to Integration. Incorporating the Digital Library into the University Research Library 35

Hermann Leskien
Der Einfluß digitaler Medien auf die bibliothekarischen Tätigkeiten . . . 51

Strategien und Kooperation bei der Digitalisierung von Archiv- und Bibliotheksgut 65

Jürgen Bunzel
Die Verteilte Digitale Forschungsbibliothek als Infrastrukturförderung der Deutschen Forschungsgemeinschaft 67

Deanna B. Marcum
Council on Library and Information Resources. Vision and Future 83

John Haeger
Better Access to Primary Sources 1983–1998 91

Elmar Mittler
Collaboratories – auf dem Weg zu neuen Formen der technisch unterstützten Kooperation 95

Organisation und Technik der Digitalisierung von Archiv- und Bibliotheksgut 101

Marianne Dörr
Planung und Durchführung von Digitalisierungsprojekten 103

Peter Exner
Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut 113

Gerald Maier
Qualität, Bearbeitung und Präsentation digitaler Bilder 129

Gerald Maier
Colormanagement bei der Farbdigitalisierung von Archivgut. Grundlagen, Hintergründe und Ausblick 179

Thomas Fricke und Gerald Maier
Automatische Texterkennung bei digitalisiertem Archiv- und Bibliotheksgut 201

Gerald Maier und Peter Exner
Wirtschaftlichkeitsüberlegungen für die Digitalisierung von Archiv- und Bibliotheksgut 223

Bereitstellung und Nutzung von digitalisiertem Archiv- und Bibliotheksgut 231

Karsten Uhde
Kontextbezogene Online-Präsentationen von Archivgut 233

Thomas Fricke
Didaktische Internet-Präsentationen von Archivgut 247

Hans-Heinrich Ebeling
Das Digitale Archiv. Ein Projekt am Stadtarchiv Duderstadt 261

Norbert Lossau
Das digitalisierte Buch im Internet. Möglichkeiten der Präsentation und Navigation 269

August Wegmann
Elektronische Bücher im Internet. Modell und standardisiertes Werkzeug zur Präsentation sequentiell abfolgender Seiten 287

Thomas Fricke
Internet-Präsentationsmodell für Zeitungen und Großformate 297

Digitalisierung im Spannungsfeld von Erhaltung und Zugänglichkeit 301

Hans Rütimann
Bestandserhaltung in einer digitalen Welt 303

Gerhard Banik
Risiken bei der Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut aus konservatorischer Sicht 311

Hartmut Weber
Langzeitspeicherung und Langzeitverfügbarkeit digitaler Konversionsformen 325

Verzeichnis der Autorinnen und Autoren 343

Vorwort

Das Authentische, heißt es bei Adorno, sei ohne Schwellenangst nicht zu haben. Diese Erfahrung haben die Archive immer wieder gemacht. Die Zurückhaltung, den Lesesaal eines Archivs zu betreten und in die authentischen Quellen Einsicht zu nehmen, hat sicher etwas mit diesem Phänomen zu tun. Das heißt für die Archive, ihre Aufgaben und Dienstleistungen transparent zu machen und an die Quellen heranzuführen, die sich in den meisten Fällen naturgemäß nicht von selbst erschließen. Die digitalen Techniken bieten hierbei vollkommen neue Möglichkeiten. Im geschützten Bereich am heimischen Bildschirm und mit der erforderlichen Zeit und Muße kann dies sicher mit weniger Streß vor sich gehen als im Lesesaal. Die Online-Präsentation von archivalischen Quellen im Kontext ist daher geeignet, neue Nutzerkreise für die Dienstleistungen der Archive zu interessieren und zugleich den Bildungsauftrag der Staatsarchive zu erfüllen.

Seit mehr als drei Jahren nutzt die staatliche Archivverwaltung das Internet, um in einem ständig wachsenden Angebot die Öffentlichkeit über die Dienstleistungen der Staatsarchive zu informieren. Nicht erst seit diesem Zeitraum beschäftigen sich die Staatsarchive und die Landesarchivdirektion mit den Folgen der Digitalisierung, sowohl im Bereich der Überlieferungsbildung als auch im Bereich der Erhaltung. So war es konsequent, aufbauend auf der Schwerpunktaufgabe Bestandserhaltung, welche die staatliche Archivverwaltung schon mehr als zehn Jahre erfolgreich wahrnimmt, die Digitalisierung im Spannungsfeld zwischen Erhaltung und Nutzung, zwischen preservation and access in bisher zwei von der Deutschen Forschungsgemeinschaft geförderten Projekten an der konservatorischen und archivfachlichen Meßlatte zu messen. Diesem Projekt verdanken wir wichtige Erkenntnisse, die in diesem Band dargestellt werden. Der Arbeit an diesem Projekt verdanken wir aber auch ein vielfältigeres und vor allem attraktiveres Online-Angebot. Die Präsentationsmodelle im Internet, die im Rahmen des DFG-Projekts erarbeitet wurden, ermöglichen selbst Laien, sich mit Hilfe der didaktischen Elemente mit Archivgut auseinanderzusetzen.

Online-Bereitstellungen von Archiv- oder Bibliotheksgut sprengen nicht nur die Mauern von Lesesälen. Eine grenzüberschreitende Zusammenarbeit erzeugt fachliche wie wirtschaftliche Synergien. Ich freue mich daher ganz besonders, daß im Rahmen des DFG-Projekts ein vielbeachtetes internationales Kolloquium in Ludwigsburg stattfand und daß dabei Archivare mit Bibliothekaren in eine intensive Fachdiskussion eintraten.

Im vorliegenden Band sind die Beiträge dieses Kolloquiums mit den Ergebnissen des Untersuchungsprojekts vereint. Der Band vermittelt damit zugleich technisches und organisatorisches Wissen, gibt Erfahrungen weiter, läßt

strategische Ansätze erkennen und enthält Anregungen zur internationalen Kooperation.

Den Autorinnen und Autoren dieses Bandes möchte ich ebenso herzlich danken wie den beiden Herausgebern Herrn Professor Dr. Hartmut Weber und Herrn Dr. Gerald Maier. Frau Eva Roll M. A. danke ich für Lektoratsarbeiten und Frau Luise Pfeifle für die Betreuung der Drucklegung. Mein besonderer Dank gilt aber auch den Förderern und Sponsoren für die finanzielle Unterstützung der internationalen Veranstaltung und des Drucks dieses Bandes, namentlich der Deutschen Forschungsgemeinschaft und den Firmen Classen-Convertronics GmbH (Essen), Herrmann & Kraemer GmbH & Co. KG (Garmisch-Partenkirchen), Heydt-Verlags-GmbH (Rottenburg am Neckar), SRZ Satz-Rechenzentrum GmbH & Co. KG (Berlin), ZfB Zentrum für Bucherhaltung GmbH (Leipzig), Zeutschel GmbH (Tübingen) und dem Staatsarchiv Zürich. Dem vorliegenden Band wünsche ich eine gute Aufnahme und hoffe, daß er dazu beitragen kann, die neuen Techniken effektiv und nutzerorientiert einzusetzen, ohne die Bestandserhaltungsaspekte zu vernachlässigen.

Stuttgart, im Mai 2000

Professor Dr. Wilfried Schöntag
Präsident der Landesarchivdirektion
Baden-Württemberg

Einleitung

Von Hartmut Weber und Gerald Maier

Nach der Erfindung des Buchdrucks mit beweglichen Lettern, die Victor Hugo als das bedeutendste Ereignis der Weltgeschichte bewertete, gibt es gute Gründe, in der Erfindung der Photographie, in der Entwicklung der Computertechnik und nicht zuletzt in der Digitalisierung weitere tiefgreifende Einschnitte zu erblicken. Auf das Archivwesen haben sich insbesondere die Möglichkeiten der Photographie und die der elektronischen Datenverarbeitung nachdrücklich ausgewirkt und von der Digitalisierung erwarten nicht Wenige revolutionäre Veränderungen.

Zu Beginn dieses Jahrhunderts feierte man den Einzug der Photographie in die Geisteswissenschaften. Mit photographischen Verfahren war die eingeschränkte Zugänglichkeit zur unikalen archivalischen Überlieferung, aber auch zu den Handschriften der Bibliotheken, durch die neuen Möglichkeiten, photographische Abbildungen anzufertigen und zu verbreiten, überwunden. In der Tendenz löste das Bestellen von Kopien, Mikrofilmen oder von Reproduktionen zunehmend das Exzerpieren im Lesesaal ab. Die Zahl der Nutzer stieg an, ihre Verweildauer aber ging zurück. In gleichem Maße stieg der Bedarf an Mikrofilmen oder Kopien an, die heute als selbstverständliche Dienstleistungen der Archive anzusehen sind oder zumindest sein sollten.

Die neuen Möglichkeiten der elektronischen Datenverarbeitung wurden vereinzelt seit den 70er Jahren, vermehrt aber seit den 80er Jahren insbesondere für die Rationalisierung und Qualitätsverbesserung der archivischen Erschließung eingesetzt. In fachlicher Hinsicht befriedigend hat sich hierbei der Wandel vom Versuch der Inhaltsrepräsentation, der den anfangs eingeschränkten Möglichkeiten der EDV Rechnung trug, zur Beschreibung und Darstellung von Entstehungsstrukturen vollzogen. Heute stehen stark erweiterte und neue Möglichkeiten der elektronischen Informationstechnik zur Verfügung, die nur mit den Schlagworten Digitalisierung und Internet-Technologie charakterisiert werden sollen. Der epochale Umbruch, den diese digitalen Technologien mit sich gebracht haben, besteht darin, daß die seit Beginn der Schriftlichkeit unveränderte Kontinuität der Verbindung von Information und Informationsträger nun aufgelöst ist. Bisher war auch die Vervielfältigung und Verbreitung unikaler Information oder der abbildbaren Erscheinungsform von Unikaten an Informationsträger wie Papier oder Film gebunden. Das ist nun nicht mehr zwingend der Fall. Digitalisierung ist zu einer Art Zauber geworden, dessen Gebrauch die Lösung vieler Probleme verheißt, die mit unikalem Archivgut verbunden waren.

Die Internet-Technologie im Sinne der weltweiten Netzinfrastruktur verbunden mit den integrierten Möglichkeiten von Hypertextanwendungen und nicht zuletzt dem Zwang zu einer gewissen Standardisierung erlaubt es, sich Informationen aller Art zu jeder Zeit auf den heimischen Bildschirm zu holen. Der Nutzer geht nicht mehr ins Archiv – das Archiv kommt zum Nutzer. In Verbindung mit diesem Netz ohne Grenzen spricht man gerne von Bibliotheken oder Archiven ohne Mauern. Über das Internet kann Bibliotheks- oder Archivgut unabhängig vom Verwahrort und von Öffnungszeiten komfortabel zugänglich gemacht werden. Dies gilt besonders auch für Archivalien, Handschriften oder sonstige Objekte, die in ihrer Erhaltung gefährdet sind und die sonst nur mit Einschränkungen unter besonderer Aufsicht benutzt werden könnten.

Mit der Digitalisierung von Archiv- und Bibliotheksgut können so zweifellos Probleme der Zugänglichkeit zu unikaler Überlieferung gelöst werden – es werden mit den neuen Techniken aber auch neue Probleme geschaffen. Wie bei der analogen Konversion der Bildinformation von Büchern oder Archivalien stellen sich die Fragen der Wiedergabequalität, der Integration in die Nutzungsumgebung, der Haltbarkeit auch im Sinne einer langfristigen Zugänglichkeit und schließlich der Wirtschaftlichkeit.

In diesem Band geht es um die Chancen, die mit dem Einsatz digitaler Techniken in den Archiven und Bibliotheken zur Absicherung ihrer Dienstleistungsqualität in der Informationsgesellschaft verbunden sind und um die Voraussetzungen und Bedingungen, unter denen die digitalen Techniken effektiv und wirtschaftlich eingesetzt werden können. Alle Beiträge sind im Rahmen eines Untersuchungsprojekts entstanden, das mit Mitteln der Deutschen Forschungsgemeinschaft in den Jahren 1997–1999 bei der Landesarchivdirektion Baden-Württemberg durchgeführt wurde. Zum einen handelt es sich um die abschließende Darstellung der Ergebnisse des Projekts durch die Mitarbeiter der Projektgruppe. Zum anderen um die Textfassungen von Referaten und Präsentationen eines Internationalen Kolloquiums, das im Rahmen des Untersuchungsprojekts vom 9. bis 11. November 1998 in Ludwigsburg veranstaltet wurde, um die in diesem Projekt gewonnenen Erkenntnisse und Erfahrungen in einen internationalen Rahmen zu stellen. Mit Zustimmung des Unterausschusses *Bestandserhaltung* der Deutschen Forschungsgemeinschaft werden die Bestandteile des Abschlußberichts in redaktionell überarbeiteter Form in die Sachgliederung dieses Bandes eingepaßt.

Zum Projekt

Das von der Deutschen Forschungsgemeinschaft geförderte Projekt *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Bücher und Archivalien* schloß an die Ergebnisse der Arbeitsgruppe *Digitalisierung* des Unterausschusses Bestandserhaltung der

Deutschen Forschungsgemeinschaft vom Oktober 1996 an, bei dem die Landesarchivdirektion ebenfalls die Federführung hatte.[1] Dieses Projekt hatte sich mit den Vor- und Nachteilen von Digitalisierungsverfahren und digitalen Konversionsformen im Vergleich zu herkömmlichen Trägermaterialien wie Papier und Mikroformen beschäftigt und Empfehlungen zur Digitalisierung von Archiv- und Bibliotheksgut erarbeitet, die im vorzustellenden Projekt erprobt und konkretisiert werden sollten.

Dem Projekt lag von vornherein ein hybrider Systemansatz zugrunde. Der in Archiven und Bibliotheken bewährte Mikrofilm sollte als Ausgangspunkt für digitale Konversionsformen und zugleich weiterhin als Langzeitspeicher dienen. Vorrangiges Projektziel war daher die Entwicklung rationeller Methoden und technischer Verfahren für die Bereitstellung von digitalisiertem Archiv- und Bibliotheksgut in Online-Medien wie dem Internet unter Berücksichtigung

- einer bedarfsgerechten Wiedergabequalität,
- einer ansprechenden und objektgerechten Präsentation,
- der Benutzerbedürfnisse,
- der technischen Realisierung und Wirtschaftlichkeit sowie
- der Langzeitverfügbarkeit.

Zur zielgerichteten Erfüllung der Projektziele wurde zu Beginn ein ausführlicher Projektstrukturplan erstellt. Der Projektstrukturplan übernahm hier zugleich die Aufgabe eines Pflichtenhefts.[2] Er ist in einzelne Arbeitspakete gegliedert, die dann jeweils in einzelne Ziele und Arbeitsschritte untergliedert sind. Als vorteilhaft hat es sich erwiesen, daß die mit dem Projekt betraute Projektgruppe sowohl mit archivfachlich als auch technisch qualifiziertem Personal besetzt war. Der Projektleiter steuerte oder moderierte zusammen mit der Projektgeschäftsstelle die Projektarbeit. Innerhalb des Projektteams erfolgte eine Aufgabenverteilung. Wichtige Aufgaben des Projektteams waren die Ausschreibung und Auswahl der notwendigen externen Dienstleister für Auftragsvergaben und die Beschaffung der erforderlichen Infrastruktur. Außerdem wurden extern vergebene Aufträge vom Projektteam intensiv begleitet.

Zunächst wurde das für die Untersuchungen benötigte Archivgut in Form von verschiedenen sogenannte *Warenkörben* zusammengestellt. Ausgewählt

---

1 Siehe dazu Marianne *Dörr*/Hartmut *Weber*, Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft. In: Zeitschrift für Bibliothekswesen und Bibliographie 44 (1997), Heft 1 und online unter: *http://www.lad-bw.de/lad/dfgdigh1.htm*.

2 Zum Projektstrukturplan siehe in diesem Band den Beitrag von Peter *Exner*, Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut, S. 113.

wurden dabei typische Vertreter einzelner Gattungen von Archivalien und Büchern, hauptsächlich aus den Beständen des Hauptstaatsarchivs Stuttgart und des Generallandesarchivs Karlsruhe. Die Objekte der Warenkörbe wurden mikroverfilmt und zum Teil vom Original direkt digitalisiert. Anschließend wurde ein Großteil der mikroverfilmten Objekte vom Film digitalisiert.

Von vornherein entschied sich die Projektgruppe für ein *Outsourcing* wichtiger Projektleistungen, um so die Kompetenz, das Know-how und technische Ressourcen anerkannter Dienstleister und Kooperationspartner in das Projekt einzubinden. Mit dem Projekt *Digitale Konversionsformen* wurden auch Erfahrungen im Bereich des sogenannten Projektmanagements gemacht, die auf andere Digitalisierungsprojekte übertragbar sind.

Die Auswertung und Weiterbearbeitung der digitalen Formen erfolgte unter folgenden Schwerpunkten:

- Digitalisierung vom Original und Digitalisierung vom Mikrofilm im Vergleich,
- Untersuchung der Digitalisierung vom Mikrofilm mit
  - Ermittlung von Qualitätsfaktoren und –kontrolle,
  - Bearbeitung und Verbesserung digitaler Bilder (*Image Enhancement*),
  - Tests mit verschiedenen Dateiformaten und Komprimierungsverfahren für die Präsentation,
- Erprobung der automatischen Texterkennung (OCR) bei der Filmdigitalisierung,
- Entwicklung von Präsentationsmodellen,
- Erarbeitung von Konzepten für die Langzeitverfügbarkeit digitaler Konversionsformen mit Prüfung des COM (Computer Output on Microfilm)-Verfahrens.

In der Summe der vielfältigen Einzelergebnisse des Projekts hat sich die Ausgangshypothese voll bestätigt: Digitale Konversionsformen von Archiv- und Bibliotheksgut aller Art ermöglichen einen attraktiven Zugang auch zu gefährdeten Objekten und stellen bei objektgerechter digitaler Bereitstellung, in Verbindung mit den erforderlichen Kontextinformationen, in aller Regel ein vollwertiges Ersatzmedium für die Nutzung der gefährdeten Originale dar.

Der unsicheren Langzeitverfügbarkeit der digitalen Konversionsformen selbst kann dadurch wirksam begegnet werden, daß die Digitalisierung konsequent vom Zwischenmedium Mikrofilm erfolgt. Die Qualitätsreserven des schwarzweißen oder farbigen Mikrofilms reichen aus, gegenwärtigen und denkbaren künftigen Ansprüchen an digitale Bereitstellungsformen vollauf zu genügen. Digitalisierungen vom Mikrofilm bestanden auch die *Nagelprobe*, welche die besonders empfindliche automatische Texterkennung (OCR) an die Qualität digitaler Bilddateien stellt. Die Haltbarkeitserwartungen von Mikrofilmen empfehlen dieses Medium als relativ anspruchslosen Lang-

zeitspeicher. So werden mit der Digitalisierung über den Film auf wirtschaftliche Weise Sicherungs- und Schutzzwecke mit dem Ziel besserer Zugänglichkeit verbunden. Wenn Digitalisierungsprojekte zugleich Ziele der Bestandserhaltung verfolgen, ist der hybride Systemansatz zwingend geboten.

Das Projekt hat zahlreiche Möglichkeiten aufgezeigt, wie man mit Standardwerkzeugen der Bildbearbeitung Rohdaten, die konsequent gewerblich vom Mikrofilm und zu Vergleichszwecken vom Original digitalisiert wurden, zu hochwertigen digitalen Mastern optimiert. Von besonderer Bedeutung ist bei farbigen Vorlagen ein effektives Colormanagement, um farbtreue digitale Master zu erzeugen. Optimale digitale Master sind die Voraussetzung für komprimierte Bilddateien, welche die Ansprüche einer Internet-Präsentation sowohl hinsichtlich der Qualität als auch der Dateigrößen erfüllen. Hierbei hat sich gezeigt, daß Bilddateien und Präsentationsmodelle der verschiedenen Gattungen von Archiv- und Bibliotheksgut optimal aufeinander abzustimmen sind. Dies gilt insbesondere bei großformatigen Vorlagen mit Detailinformationen geringer Größe wie Karten, Pläne oder Zeitungen. Im Projekt gelang es, solche nachnutzbaren Präsentationsmodelle für eine Fülle von Gattungen zu entwickeln. An diesen Entwicklungen konnte auch der Aufwand für die objektgerechte Bildbearbeitung und Bereitstellung digitaler Konversionsformen von Archiv- und Bibliotheksgut abgeschätzt werden, der die Kosten für die Digitalisierung je nach dem Mehrwert, der mit der Präsentation verbunden wird, um ein vielfaches übertrifft.

Die digitale Bereitstellung von Archiv- und Bibliotheksgut ist zugleich ein Problem, ein Massenproblem, da sie in der Regel mit einer jeweils großen Zahl von Seiten verbunden ist. Es gilt daher, die Arbeitsabläufe zu rationalisieren, um sie möglichst kostengünstig zu gestalten. Für die Bildbearbeitung und digitale Bereitstellung müssen daher nachnutzbare Werkzeuge entwickelt und diese in einem optimierten und standardisierten Arbeitsablauf (*workflow*) integriert werden. Dies war nicht Ziel des nun abgeschlossenen Projekts. Im Rahmen dieses Projekts konnten aber insbesondere für die jeweils objektgerechte Bildbearbeitung vom Mikrofilm oder unmittelbar digitalisierter Vorlagen automatisierte Verarbeitungsabläufe (*Makros*) entwickelt werden. Prinzipversuche zum Colormanagement haben gezeigt, daß sich auch in diesem Bereich mit zusätzlichem Aufwand ein durchgängiger standardisierter Ablauf mit automatisierten Werkzeugen entwickeln läßt. Schließlich beweist der mit relativ preiswerter Standardsoftware entwickelte Prototyp eines automatisierten Workflows für ein Präsentationsmodell, mit dem Bücher und andere sequentiell organisierte Objekte digital bereitgestellt werden können, daß ähnliche Werkzeuge und standardisierte Abläufe sich auch für die vorliegenden Präsentationsmodelle entwickeln lassen, mit deren Hilfe dann andere Gattungen von digitalisiertem Archiv- und Bibliotheksgut rationell vorbereitet und im Internet bereitgestellt werden können. Solche Werkzeuge und Empfehlungen zum optimalen Arbeitsablauf werden andere Archive und Bibliotheken in die Lage versetzen, mit relativ geringem Aufwand Archiv- und Bibliotheksgut

digital bereitzustellen, und tragen so zur Verbreiterung und Qualitätssicherung der *Verteilten digitalen Forschungsbibliothek* bei. So war es nur konsequent, mit weiterer Unterstützung durch die Deutsche Forschungsgemeinschaft ein Folgeprojekt in diese Richtung in Angriff zu nehmen.

Die Ergebnisse dieses Projekts im engeren Sinne werden durch die Beiträge von Peter Exner, Thomas Fricke, Gerald Maier, Karsten Uhde, Hartmut Weber und August Wegmann dargestellt.

Zu den Beiträgen dieses Bandes

Ein erster Themenbereich beschäftigt sich mit *Fachkonzepten für digitale Archive und Bibliotheken*. Sönke Lorenz stellt aus der Sicht des Historikers die *Anforderungen der Geschichtsforschung an einen digitalen Arbeitsplatz* vor und unterstreicht damit, daß sich Digitalisierungsprojekte konsequent am Bedarf orientieren sollten. Die *Auswirkungen der Digitalisierung auf die archivischen Fachaufgaben* beschreibt Angelika Menne-Haritz. Sie verdeutlicht die neuen Möglichkeiten der Archive als Anbieter von Informationspotentialen über den *Digitalen Lesesaal*, *virtuelle Magazine und Online-Findbücher* im Sinne einer Kundenorientierung. Carol Mandel erläutert die Integration und den Stellenwert digitaler Dokumente innerhalb der Struktur und Organisation einer großen amerikanischen Universitätsbibliothek. Ein Beitrag über den *Einfluß digitaler Medien auf die bibliothekarischen Tätigkeiten* und die Organisation einer großen Bibliothek von Hermann Leskien vervollständigt den ersten Themenbereich.

Strategien und Kooperation bei der Digitalisierung von Archiv- und Bibliotheksgut werden in einem zweiten Themenbereich vorgestellt. Die Beiträge beschäftigen sich vor allem mit den verschiedenen Formen und Möglichkeiten einer internationalen Kooperation zwischen Archiv- und Bibliothekseinrichtungen auf dem Gebiet der Digitalisierung von Archiv- und Bibliotheksgut. Über die Bedeutung der *Verteilten Digitalen Forschungsbibliothek als Infrastrukturförderung der Deutschen Forschungsgemeinschaft* handelt der Beitrag von Jürgen Bunzel. Deanna B. Marcum stellt die Entstehung, Aufgabenfelder und Ziele des *Council on Library and Information Resources* vor. Die Einrichtung hat in den USA die Bedeutung einer Koordinationsstelle verschiedener Aktivitäten auf dem Gebiet der Bestandserhaltung und der digitalen Bibliotheken. Der Beitrag von John Haeger, von der weltweit tätigen Research Library Group, thematisiert die Integration archivischer und bibliothekarischer Findmittel unter einer gemeinsamen Benutzeroberfläche und unter Verwendung der Dokumentenbeschreibungssprache EAD (Encoded Archival Description). Elmar Mittler behandelt in seinem Beitrag verschiedene Möglichkeiten einer zukünftigen Zusammenarbeit zwischen Forschung und Bibliotheken in den USA und Deutschland.

Mit *Organisation und Technik der Digitalisierung von Archiv- und Bibliotheksgut* beschäftigt sich der dritte Themenbereich. Den Anfang macht ein Beitrag von Marianne Dörr über die *Planung und Durchführung von Digitalisierungsprojekten* aufgrund der Erfahrungen des Münchner Digitalisierungszentrums. Der Beitrag von Peter Exner über die *Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut* stellt die in diesen Bereichen im Rahmen des Projekts *Digitale Konversionsformen* gemachten Erfahrungen und Ergebnisse vor. Ebenfalls in diesen Kontext gehört der Beitrag von Gerald Maier über *Qualität, Bearbeitung und Präsentation digitaler Bilder*. Vorgestellt werden hier unter anderem Methoden der Qualitätssicherung und Bildnachbearbeitung und die Verwendung geeigneter Dateiformate für die Präsentation digitaler Bilder in Online-Medien wie dem Internet. In diesem Zusammenhang steht auch ein weiterer Beitrag desselben Autors über das Thema *Colormanagement bei der Farbdigitalisierung von Archivgut*. Über die im Rahmen des Projekts *Digitale Konversionsformen* angestellten Untersuchungen zum Einsatz der *automatischen Texterkennung bei digitalisiertem Archiv- und Bibliotheksgut* berichten Thomas Fricke und Gerald Maier. Schließlich werden *Wirtschaftlichkeitsüberlegungen für die Digitalisierung von Archiv- und Bibliotheksgut* in einem Beitrag von Gerald Maier und Peter Exner thematisiert.

Ein vierter Bereich beschäftigt sich mit der *Bereitstellung und Nutzung von digitalisiertem Archiv- und Bibliotheksgut*. Vorgestellt werden hier verschiedene Möglichkeiten und Modelle der Internet-Präsentation von digitalisiertem Archiv- und Bibliotheksgut. Karsten Uhde stellt in seinem Beitrag *kontextbezogene Online-Präsentationen von Archivgut* vor, die im Rahmen des vorgestellten Projekts entwickelt wurden. Ebenfalls in den Kontext des Projekts gehört der Beitrag von Thomas Fricke über *didaktische Internet-Präsentationen von Archivgut*. Hans-Heinrich Ebeling beschreibt in seinem Beitrag *das Digitale Archiv* Duderstadt. Bei diesem gemeinsamen Projekt des Duderstädter Stadtarchivs mit dem Max-Planck-Institut für Geschichte in Göttingen wurden Teile der älteren Bestände des Stadtarchivs digitalisiert. Ein breites Spektrum von Möglichkeiten der Präsentation *des digitalisierten Buchs im Internet* und Möglichkeiten der Navigation in Büchern sind Thema des Beitrags von Norbert Lossau. Modell und standardisiertes Werkzeug zur Präsentation sequentiell abfolgender Seiten beschreibt August Wegmann für *Elektronische Bücher im Internet*. Der vierte Bereich wird abgerundet durch einen Beitrag von Thomas Fricke über ein *Internet-Präsentationsmodell für Zeitungen und Großformate.*

Die *Digitalisierung im Spannungsfeld von Erhaltung und Zugänglichkeit* wird in einem letzten Bereich thematisiert. Hans Rütimann beleuchtet in seinem Beitrag die Problematik der *Bestandserhaltung in einer digitalen Welt*. Auf mögliche *Risiken bei der Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut aus konservatorischer Sicht* macht Gerhard Banik in seinem Beitrag aufmerksam. Der Themenbereich wird von Hartmut Weber abgeschlossen mit einem Beitrag zur *Langzeitspeicherung und Langzeitverfügbar-*

keit digitaler Konversionsformen. Neben der Vorstellung verschiedener Konzepte wie Migration, Konversion und Emulation wurden auch Ergebnisse einer im Rahmen des Projekts *Digitale Konversionsformen* durchgeführten Testreihe mit dem COM (Computer Output on Microfilm)-Verfahren präsentiert.

Fachkonzepte für digitale Archive und Bibliotheken

Anforderungen der historischen Forschung an einen digitalen Arbeitsplatz

Von Sönke Lorenz

Die von der Landesarchivdirektion im Internet vorgestellte Konzeption zum Projekt *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Archivalien und Bücher* beginnt mit einer Feststellung, der sich ein Fragesatz anschließt.

Ich darf zitieren: *Angesichts des drohenden unwiederbringlichen Verlustes von Kulturgütern wird die digitale Technologie zunehmend als Problemlösung angeboten. Liefert die Digitalisierung vom Zerfall bedrohter Archivalien und Bücher tatsächlich einen sicheren Langzeitspeicher, oder stellt sie vielmehr ein attraktives Zugriffs- und Nutzungsmedium dar, auf das der Benutzer per Mausklick bei seinen Recherchen in Archiven und Bibliotheken zurückgreift?*[1]

Zu der Frage, ob und wie die Digitalisierung des bedrohten Kulturgutes einen sicheren Langzeitspeicher liefert, möchte ich mich nicht äußern, da dieses Thema an anderer Stelle kompetent behandelt wird.[2] Ich möchte hier nur einige Überlegungen aus der Sicht des Benutzers vortragen, der irgendwann per Mausklick auf das digitalisierte Kulturgut zurückgreifen möchte. Denn – und darin liegt mir eine Schwäche der eingangs zitierten Formulierung – es geht nicht um ein alternatives *oder*, sondern um ein kumulatives *und*. Die Digitalisierung wird doch sicher nicht nur deshalb betrieben, weil beispielsweise Papierfäule und Tintenfraß Archivalien vernichten, sondern weil die Benutzung oft von unvermeidlichen Schäden begleitet ist.

Betrachten wir als Beispiel die Württembergischen Urkunden bis 1300 des Hauptstaatsarchivs Stuttgart – Pergament und Siegel sind aufs äußerste gefährdet.[3]

So denke ich, eine Digitalisierung sollte unbedingt auch von einer Aufbereitung für den Benutzer begleitet sein, um die gefährdeten Archivalien, Handschriften, Inkunabeln, Frühdrucke und sonstiges auf einem Wege der

---

1 Siehe *http://www.lad-bw.de/digpro.*

2 Siehe dazu in diesem Band den Beitrag von Hartmut *Weber:* Langzeitspeicherung und Langzeitverfügbarkeit digitaler Konversionsformen. S. 325.

3 HStAS Bestand A 601. Vgl. Übersicht über die Bestände des Hauptstaatsarchivs Stuttgart. Altwürttembergisches Archiv (A-Bestände), bearbeitet von Hans-Martin Maurer (Veröffentlichungen der Staatlichen Archivverwaltung Baden-Württemberg, Bd. 32). Stuttgart 1975. S. 194.

Wissenschaft zugänglich zu machen, welche den an die angeführten Quellen gerichteten Fragestellungen in den meisten Fällen Genüge leistet. Selbstverständlich wird beispielsweise einem Editor von Urkunden weiterhin die Autopsie möglich bleiben können.

Wie aber hat man sich eine solche Aufbereitung vorzustellen? Anders gefragt, was haben die für die Digitalisierung verantwortlichen Institutionen zu bedenken, um den Benutzerbedürfnissen Rechnung zu tragen? Wenn ich mich im folgenden auf die Archive und Handschriftenabteilungen der Bibliotheken konzentriere, dann hat das gleich mehrere gewichtige Gründe, von denen kurz auf die zwei wichtigsten verwiesen sei. Zum einen geht es hier anders als bei den normalen Bibliotheksbeständen in aller Regel um Quellen, deren Unikatcharakter bisher lediglich erlaubte, sie in den Lesesälen zu benutzen. Zum andern handelt es sich in den meisten Fällen nicht um gedruckte, sondern um handschriftliche Quellen, die sich nur dem erschließen, der über mehr oder weniger gediegene Fähigkeiten in Paläographie sowie Akten- und Urkundenlehre verfügt, um nur diese beiden zentralen hilfswissenschaftlichen Fächer aufzuzählen.

Nun gibt es bisher ja bekanntlich bereits eine ganze Reihe von möglichen Vorgehensweisen, sich solchen Archiv- und Bibliotheksbeständen zu nähern. Ich greife kurz zwei mir besonders üblich erscheinende heraus. Entweder hat man über die Literatur Kenntnis gewonnen, daß im Archiv X die interessierende Handschrift Y unter der Signatur Z verwahrt wird. Oder man ist an einem bestimmten Thema interessiert, von dem man annimmt, das Archiv X verfüge ob einer gewissen Zuständigkeit über einen Bestand Y, in dem sich einschlägiges Quellenmaterial verbergen könnte. In beiden Fällen wäre es sehr hilfreich, wenn das Archiv über Beständeübersichten verfügte, auf die der Benutzer im Internet online zugreifen könnte. Bekanntlich hat das Universitätsarchiv Tübingen vor anderthalb Jahren seine Beständeübersicht ins Internet gestellt.[4] Hier sei zudem noch auf das Projekt *Nordrhein-Westfälische Archive im Internet* hingewiesen, bei dem im Staatsarchiv Münster ein Internet-Online-System für Beständeübersichten entwickelt wird.[5]

Ich halte es für eine der wichtigsten *Anforderungen der historischen Forschung an einen digitalen Arbeitsplatz*, daß die Archive solche Beständeübersichten ins Internet stellen. Vielleicht wird es dabei auch möglich sein, den vorhandenen Übersichten eine neue Struktur zu geben, die dem EDV-Benutzer die Recherche erleichtert. Aber wie dem auch sei, die Ebene der Beständeübersichten erscheint das vorrangigste Ziel, um das Online-Angebot im Internet sinnvoll aufzubauen.

Als zweite Ebene bieten sich die Findbücher an. Bisher bekanntlich vorrangig oder ausschließlich im Lesesaal benutzbar, wäre es sicher eine große

---

[4] URL: *http://www.uni-tuebingen.de/UAT.*

[5] URL: *http://www.archive.nrw.de.*

Erleichterung für die Recherche, wenn man die Durchsicht der Findbücher über das Internet vornehmen könnte. So wurde an der Archivschule in Marburg ein online-fähiges Findbuch erarbeitet.[6] Das Tübinger Universitätsarchiv hat am 1. September 1998 sein erstes Repertorium ins Internet gestellt, mittlerweile sind es bereits fünf Repertorien. Man braucht nicht zu betonen, daß mit Hilfe solcher Findbücher die Recherche – im wahrsten Sinne des Wortes – praktisch aus dem Lesesaal an den heimischen Bildschirm verlagert und damit *zeit- und ortsunabhängig* wird. Vielleicht noch wichtiger sind die Möglichkeiten der EDV, die einen Qualitätssprung hinsichtlich einer umfassenden Bestandsdurchdringung durch den Benutzer erlauben. So werden nicht nur Archivalien vor einer fruchtlosen und unergiebigen Benutzung bewahrt, sondern neue Maßstäbe für die wissenschaftliche Erschließung gesetzt, die dem Historiker das Handwerk wesentlich erleichtern.

Beständeübersichten und Findbücher, die online ins Internet gestellt werden, erlauben es beispielsweise auch, die provenienzorientierte Tektonik der Archive virtuell zu verändern und an die in der Regel stärker pertinenzorientierten Fragestellungen der Wissenschaft anzupassen. Mit Hilfe von Sach- und Lokalpertinenzen sowie einem Personenindex wird ein mehrdimensionaler Zugriff möglich, der in jeder Beziehung einen Qualitätssprung darstellt. Mit Hilfe eines Katasters der in ihrer jeweiligen historischen Dimension zu erarbeitenden Territorien und Schriftgutproduzenten – wie beispielsweise Behörden, Gerichte, öffentliche und private Einrichtungen – ließe sich für Forschungszwecke der Verbleib des Archivguts auch archivübergreifend und gezielt nachweisen.

Die dritte Ebene schließlich gehört der Archivalie, dem einzelnen Objekt, das online im Internet und/oder offline in Form von CD-ROMs verfügbar gemacht wird, wie vor kurzem am Beispiel der Revolutionskartei exemplarisch vorgeführt.[7] Auf dieser Ebene sind vorrangig Corpora ausgewählter, zentraler und häufig genutzter archivalischer Quellen in digitalisierter Form bereitzustellen. Ich denke beispielsweise an den Bestand A 602 im Hauptstaatsarchiv Stuttgart, die sog. Württembergischen Regesten (1301–1500).[8] Dieser Bestand, der seit langem im Zentrum wissenschaftlicher Bemühungen steht, sollte schon aus konservatorischen Gründen unbedingt und rasch einer digitalen Erfassung zugeführt werden.

---

6 URL: *http://www.archivschule.de*.

7 Siehe dazu die Datenbank auf CD-ROM mit Recherchemodul in Heinrich *Raab:* Revolutionäre in Baden 1848/49. Biographisches Inventar für die Quellen im Generallandesarchiv Karlsruhe und im Staatsarchiv Freiburg, bearbeitet von Alexander Mohr (= Veröffentlichungen der Staatlichen Archivverwaltung Baden-Württemberg, Bd. 48). Stuttgart 1998.

8 Vgl. Übersicht über die Bestände des Hauptstaatsarchivs Stuttgart. Altwürttembergisches Archiv (A-Bestände), S. 194.

Wichtig – um nicht zu sagen entscheidend – ist für diese Ebene die Qualität der Wiedergabe am Bildschirm und über den Drucker. Hier kommt es unbedingt auf eine einwandfreie Lesbarkeit der Archivalie an, die dem Benutzer zudem Vergleiche mit anderen Stücken erlauben muß, um beispielsweise nach Schreiberhänden sortieren zu können. Zu dieser Ebene wäre hier sicher noch einiges näher auszuführen, doch möchte ich lieber noch kurz zwei weitere Ebenen ansprechen, die den Benutzer aktiv *ins Spiel bringen*.

Wir erleben immer wieder, daß große wissenschaftliche Vorhaben, die der Edition und/oder Regestrierung von Quellen dienen, sich zu Jahrhundertprojekten ausweiten. Da ein an den Ergebnissen solcher Unternehmen interessierter Wissenschaftler wenigstens heute noch nicht in der Lage ist, sein Lebensalter beispielsweise durch Tiefschlaf oder andere Mittel so zu verlängern, daß er schließlich doch noch das Erscheinen miterleben kann, bietet sich die EDV geradezu als ein idealer Mittler an. Ich denke, in vielen Fällen lassen sich Einzel- und Zwischenergebnisse ins Internet geben, mit deren Hilfe dann andere bereits weiterarbeiten können. Hier ist die historische Forschung gefordert, denn sie macht noch viel zu wenig von den neuen Möglichkeiten Gebrauch.

Die letzte Ebene schließlich, also die fünfte, gehört solchen Vorhaben, die von der Forschung speziell für Online- oder Offline-Recherchen erstellt werden, wie beispielsweise die Patrologia latina von Migne.[9] Statt in über 50 Bänden zu blättern und in jahrelanger Arbeit Stichworte zu sammeln, lassen sich heute in wenigen Minuten Recherchen von höchster Präzision und Vollständigkeit durchführen. Gerade die Edition von Urkunden wird künftig nicht mehr am Internet vorbeikommen, oder – um es positiv zu formulieren – erst die EDV ist in der Lage, die Vorteile einer kritischen Edition voll auszuschöpfen, die ja eine kaum noch überschaubare Vielfalt möglicher Fragestellungen und Auswertungen bereithält.

---

[9] Siehe beispielsweise im Angebot der CD-ROM-Datenbanken der Universitätsbibliothek Tübingen (*http://www.uni-tuebingen.de/ub/cdrom/pld.htm*).

Digitaler Lesesaal, virtuelle Magazine und Online-Findbücher
Auswirkungen der Digitalisierung auf die archivischen Fachaufgaben

Von Angelika Menne-Haritz

Für Archive hat die zunehmende Verbreitung und Vereinfachung der Nutzung elektronischer Instrumente zweifache Wirkungen. Einerseits verändern sich die Möglichkeiten, Instrumente zur Präsentation und Öffnung ihrer Bestände herzustellen. Das hat Einfluß auf die Diskussionen um Erschließungsmethoden und -techniken.[1] Zum anderen wird aber auch der Rohstoff der Archivierung, die Unterlagen aus Verwaltungstätigkeit, verändert und das fordert neue Reaktionsweisen heraus.[2]

Ich möchte mich zunächst mit dem zweiten Aspekt beschäftigen, um anschließend auf die archivischen Arbeitsverfahren einzugehen. Eine deutlich zu spürende Auswirkung der neuen Entwicklungen ist nämlich selbst in der Intensivierung der Wahrnehmung von Verwaltung durch die Archive zu erkennen. Je vielfältiger die Erscheinungsformen der Verwaltungsaufzeichnungen sind, um so deutlicher wird die Anforderung einer gründlichen Analyse vor Beginn einer archivischen Bearbeitung. Und so entsteht eine zentrale Herausforderung für die archivischen Tätigkeiten darin, ihre Analysekompetenz zu vertiefen und ihr Instrumentarium zu diesem Zweck auszubauen.

---

1 Vgl. die Entwicklung der ISAD (G) und die Diskussion um die Encoded Archival Description im American Archivist, vol. 60 und 61, 1997 sowie Mechthild *Black-Veldtrupp*: Findbücher im Internet – Möglichkeiten ihrer Präsentation. In: Archivische Erschließung: Methodische Aspekte einer Fachkompetenz. Beiträge des 3. Archivwissenschaftlichen Kolloquiums der Archivschule Marburg. Hg. von Angelika *Menne-Haritz*. Marburg 1999. S. 123–138.

2 Beide Aspekte hängen eng miteinander zusammen, erfordern jedoch getrennte Strategien. Wie schwer allerdings die Abgrenzung von einander ist zeigt die häufige Umorganisation und Neuformierung der entsprechenden Komitees des ICA, der zunächst ein Komitee für Automation hatte, das 1992 aufgeteilt wurde und nun Descriptive Standards und Current Records behandelt. Zum Stand der Diskussion vgl. die Proceedings from the Working Meeting on Electronic Records, Pittsburgh, PA, May 1997. In: Archives and Museums Informatics, vol. 11 (1997) nos. 3–4.

Beratung für die Verwaltung

Archive benötigt man, wenn man mit Schriftlichkeit operiert. Da bisherige, analoge schriftliche Aufzeichnungen die Alterungseigenschaften ihrer Trägermaterialen annahmen, konnte man ihre Kenntnisnahme sowie die Retentionsfähigkeit der Mitteilungen über Operationen steuern, die auf den Träger bezogen waren. Man schickt einen Brief, also ein Ding, an eine Person, und geht davon aus, daß sie die Nachricht zur Kenntnis nimmt. Man verschließt den Brief zuvor und darf annehmen, daß er nicht unbefugt geöffnet wird. Denn jede Öffnung erfordert eine Manipulation des Dinges, die nur schwer ohne Spuren zu hinterlassen, vorzunehmen ist. Was analog aufgezeichnet ist, verbindet sich so vollständig mit dem Material, daß die logische Trennung zwischen beidem erst mit der Verselbständigung der Nachricht in elektronischer Form denkbar wurde. Bisheriger Umgang mit Schriftlichkeit implizierte also immer den stellvertretenden Umgang mit den Aufzeichnungsmaterialien. Eine vorwiegend mündliche Kultur hat andere Formen für die gleichen Zwecke. So bewirkt sie etwa die Retention durch ständige Wiederholung, die ihrerseits durch Formelhaftigkeit der Aussage erleichtert wird. Deshalb legt sich eine mündliche Aussage viel weniger fest als eine schriftliche Aufzeichnung. Sie kann aber auch nicht die Präzision der Schriftlichkeit erreichen. Andererseits integriert sie ihre Zuhörer stärker und direkter als die zeit- und ortsversetzten, von einander getrennten Operationen des Schreibens und Lesens.[3]

Dieser Gegensatz mündlicher und schriftlicher Kommunikation, der die letzten 2 000 Jahre unserer kulturellen Entwicklung prägte, wird nun durch ein neues Aufzeichnungsmedium aus dem Gleichgewicht gebracht, in dem sowohl Elemente schriftlicher Aufzeichnungen wie mündlicher Kommunikation zu erkennen sind. Der kanadische Archivar Hugh Taylor spricht von dem neuen Phänomen konzeptioneller Mündlichkeit, das einen Paradigmenwechsel der Archivwissenschaft nötig mache.[4] Er verweist damit auf die Flüchtigkeit im Gegensatz zur gewohnten Stabilität von schriftlichen Aufzeichnungen in elektronischer Form: Wo bisher eine Entscheidung erforderlich war, wenn eine Aufzeichnung nicht aufbewahrt werden sollte, ist nun im Gegensatz dazu eine Entscheidung und ein Komplex von speziellen Vorkehrungen erforderlich, wenn sie erhalten bleiben soll. Doch es geht nicht allein darum, daß die Entscheidungsanforderung umgekehrt wird. Darüber hinaus werden zusätzliche Entscheidungen und Operationen nötig, wo man bisher einen Automatismus gewohnt war. Zusätzliche Operationen benötigen aber Zeit und personelle wie materielle Ressourcen.

---

[3] Eric A. *Havelock:* Schriftlichkeit. Das griechische Alhabet als kulturelle Revolution. Weinheim 1990.

[4] Hugh A. *Taylor:* Transformation in the Archives: Technological Adjustement or Paradigm Shift? In: Archivaria 25 (Winter 1987–88) S. 12–28.

Das stellt ein grundlegendes Prinzip der Existenz von Archiven in Frage. Bisher entstanden ihre Bestände aus den Aufzeichnungen, die auf Grund einer nicht getroffenen Entscheidung über die Vernichtung oder Erhaltung in den Behörden zur Aussonderung anstanden. Archive entlasteten die Behörden nicht nur von der Aufgabe der Erhaltung, sondern, was noch mehr zählt, weil es Kosten und Zeit einspart, von dem Zwang zur Auswahl und zur Entscheidung über Aufbewahrung oder Vernichtung. Sollen die Archive also nun in die Produktion ihrer Rohstoffe eingreifen, sollen sie ihre Materialien sozusagen bestellen? Können sie von den Behörden verlangen, zusätzliche Entscheidungen zu fällen und Operationen zu unternehmen, die nur den Zweck haben, dem Archiv adäquates Material zu liefern, dafür aber Zeit und Mittel in Anspruch nehmen, die die Behörde für ihre Aufgabenerledigung einsetzen könnte? Und führt diese Anforderung nicht eventuell sogar dazu, den Sinn der Archive in einer neuen Umgebung in Frage zu stellen, wenn sie früher ein Instrument zur Problemlösung und Entlastung waren, nun aber selbst Probleme aufwerfen und neue Belastungen verursachen?

Die neuen Anforderungen der Archive an die Verwaltungen werden in der archivwissenschaftlichen Diskussion unter den Begriffen des Life Cycle[5] oder des Records Continuum[6] artikuliert. Dabei werden Regeln erarbeitet, wie etwa ein Australischer Normentwurf für Schriftgutverwaltung,[7] nach denen sich der Umgang mit elektronischen Aufzeichnungen in der Verwaltung zu richten haben soll, und die Anforderungen der archivischen Auswertung von Beginn an berücksichtigen. Demnach soll etwa bei jeder in einer Verwaltung getroffenen Entscheidung eine Aufzeichnung erstellt und für die spätere Archivierung festgehalten werden, welche Gründe für oder gegen die Entscheidung sprachen und warum sie getroffen wurde. Damit soll bei der Kommunikation bedacht werden, daß sie in eine archivfähige Form gebracht wird und daß sie zudem selbst ihre Archivwürdigkeit begründet. Beides aber sind Ana-

---

5 Es wird oft vergessen, daß der Begriff des Life Cycle um einiges älter ist als die Diskussion um die electronic records. Vgl. dazu Terry *Cook:* What is Past is Prolog: A History of Arcival Ideas Since 1898, and the Future Paradigme Shift. In: Archivaria 43 (1977) S. 17–61.

6 Auch dieser Begriff ist älter und stammt aus dem australischen Archivwesen. Er richtet sich gegen die Unterteilung des Life Cycle in verschiedene Phasen und will statt dessen eine kontinuierliche Entwicklung sehen. Vgl. Jay *Atherton:* From Life Cycle to Continuum: Some Thoughts on the Records Management – Archives Relationship. In: Archivaria 21 (1985–86) S. 17–61 sowie die Neubelebung des Konzepts mit einem erweiterten Anspruch: Sue *McKemmish* and Michael *Pigott*, Records Continuum: Ian Maclean and Australian Archives First Fifty Years, Clayton 1994. – Das Konzept des Continuum richtet sich ebenso wie das Life Cycle Konzept ausdrücklich gegen die Unterscheidung eines Verwaltungszwecks vom Archivierungszweck und sieht das Records Management als archivische Aufgabe an.

7 Standard Australia Committee IT/21 on Records Management Systems.

lysekriterien. Wenn sie als Produktionsrichtlinien verwendet werden, machen sie trotzdem die nachträgliche Analyse nicht überflüssig. Sie machen sie nur komplizierter. Es liegt auf der Hand, daß – neben der Lahmlegung jeder Behörde durch solche Anforderungen – auch das eigentliche Ziel, nämlich Begründungen zu bekommen, verfehlt wird. Denn sobald in der Verwaltung dem Archiv bewußt zugearbeitet wird, verliert das Archiv seinen Sinn. Nur wenn es der Verwaltung weiterhin Entscheidungen abnehmen kann, also sich nicht allein auf die Verwahrung beschränkt, kann es eine sinnvolle Dienstleitung anbieten.

Nicht die Aufstellung von Forderungskatalogen mit der Konsequenz zusätzlicher Belastungen wird die Nützlichkeit der Archive sichern. Statt dessen wird Beratung benötigt, die bei der Erledigung der eigenen Aufgaben hilft und den Entscheidungsdruck verringert. Diese Leistung ist unabhängig vom Kommunikationsmedium, das in der Verwaltung verwendet wird. Ja sie kann sogar den Einsatz der Kommunikationsmedien selbst zum Gegenstand machen, ist aber nur dann sinnvoll und nachhaltig, wenn sie ausschließlich die eigenen Zwecke der Verwaltung, die Verständlichkeit ihrer Kommunikation und die Ausschaltung von Mißverständnissen zum Ziel hat. Genau hier, bei der Nutzung von Aufzeichnungsformen und -medien liegt ein hoher Beratungsbedarf der Verwaltung. Denn mit den elektronischen Medien hat sich der Gestaltungsspielraum bei der Nutzung von Kommunikationsformen vervielfacht. Es ist zwar möglich, solche Kommunikationsinstrumente einzusetzen, die den beabsichtigten Zweck am besten fördern. Doch ihre jeweilige Wirkungsweise und Funktionalität für die aktuellen Entscheidungserfordernisse sind nur noch mit professioneller Unterstützung erkennbar. Verantwortungsbewußte Entscheidungen brauchen deshalb Kriterienkataloge und vorformulierte Instrumentarien. Sie lassen sich aus der in den Archiven versammelten Erfahrungen der Verwaltung mit mündlichen und schriftlichen Kommunikationsformen erarbeiten.

Weder Life-Cycle noch Records Continuum sind Konzepte, die eine archivische Beratung der Verwaltungsarbeit begründen können. Sie setzen die Kontinuität der Zwecke während der Erstellung, der Nutzung und der zeitweiligen oder endgültigen Aufbewahrung von Aufzeichnungen voraus. Deshalb können sie die Gründe für die Entstehung der Aufzeichnungen, etwa anstelle von mündlicher Kommunikation nicht erkennen und arbeiten mit Vorstellungen, die die dingliche Einheit analoger Aufzeichnungen schlicht elektronisieren. Denn diese Konzepte können sich keine funktionalen Äquivalente in elektronischer Form vorstellen, weil sie die Existenz der Aufzeichnungen bereits voraussetzen und ihre Entstehung nicht mehr selbst beobachten können. Sie können sie nur als entstanden akzeptieren.

Das ist einer der Aspekte des von Hugh Taylor geforderten Paradigmenwechsels. Es werden nicht nur und nicht in erster Linie neue technische Kompetenzen erforderlich. Wichtiger sind neue Beratungs- und Kooperationskompetenzen, die sich nicht darauf beschränken dürfen, Anforderung zu

formulieren, sondern die der Verwaltung helfen, neue Handlungsmöglichkeiten zu entwickeln. Dazu muß der archivische Blick über die Formen, über die Analyse des Layouts von Schriftstücken, über die Komposition von Akten[8] hinaus deren Funktionen in der jeweiligen Arbeitsumgebung, in der sie entstanden sind, umfassen und verstehen. Damit erwirbt der Beruf die Kompetenzen, die sein gewohntes Problemlösungspotential erhalten können und die seine Unverzichtbarkeit für eine leistungsfähige Verwaltung erneut begründen.

Die Öffnung der Bestände

Der andere Aspekt betrifft die Öffnung der aus der Verwaltung übernommenen Aufzeichnungen für die Nutzung, die nach den Archivgesetzen jedem ohne Ausnahme, allerdings unter Abwägung von durch die Öffnung eventuell beeinträchtigten anderen Rechten, zusteht.

Neue Informationstechnologien waren seit ihrer Verfügbarkeit interessant für Archive. In den 70er Jahren versuchte man die Sortier- und Recherchierfähigkeiten von Datenbanken für Stichwort- und thesaurusgestütztes Retrieval einzusetzen. In den 80er Jahren entstanden nach heftigen Diskussionen über den Sinn solcher Verfahren mit AIDA in Hannover und MIDOSA in Stuttgart Systeme, die nicht mehr mit Stichworten, sondern mit Titeln von Verzeichnungseinheiten operierten. Das war ein wichtiger Schritt vom Versuch der Inhaltsrepräsentation zur Darstellung von Entstehungsstrukturen. Doch konnten nur die Verarbeitungsvorgänge elektronisch erledigt werden. Die Darstellung selbst erfolgte durch einen Ausdruck auf Papier, was generell bei den beteiligten Informatikern zu höchstem Erstaunen führte, sich aber in der Praxis der Archive durchgesetzt hat. Alle heute verfügbare Erschließungssoftware besitzt die Funktionalität zum Ausdruck von Findbüchern.

Der logisch folgende, nächste Schritt war erst mit der Internettechnologie möglich. Denn sie macht mit ihrer navigierenden, assoziativen und selbstbestimmten Nutzungsweise ein elektronisches, funktionales Äquivalent zu

---

8 Akten werden nicht für die Aufbewahrung geschaffen, sondern für die verbrauchende Nutzung bei der Erarbeitung von Entscheidungen. Ihr Informationsgehalt ist üblicherweise gering und zudem durch den Verwendungszweck gefiltert. Doch ihr Informationspotential ist nicht eingrenzbar und nicht eindeutig benennbar. Denn die Aufzeichnungen, die sie enthalten, entstehen für die Entscheidungsproduktion, nicht die verläßliche Erinnerungsfähigkeit der Behörde, noch ihre Rechenschaftsfähigkeit oder die Kontinuität ihrer Entscheidungen. Diese Phänomene können nicht als solche direkt angestrebt werden, weil sie, wenn sie ins Zentrum der Aufmerksamkeit gezogen werden, automatisch die Problemlösungskapazität okkupieren und von den Sachen ablenken. Es ist so ähnlich wie mit der Brille. Wenn man sie sieht, etwa weil ein Staubkorn oder Fingerabdruck stört, sieht man die mit ihrer Hilfe betrachtete Umwelt solange nicht mehr.

Findbüchern möglich. Das etwa sollte mit dem Online-Findbuch der Archivschule Marburg demonstriert werden. Es wurde parallel zu der Encoded Archival Description (EAD) in den USA entwickelt und zeigt zahlreiche Ähnlichkeiten, soweit von hier aus erkennbar, aber auch einige prinzipielle Unterschiede. Die Internet-Technologie eröffnet weitere, über die Funktionen des Findbuchs hinausgehende Möglichkeiten und schafft so Raum für neue Wege zur Verbesserung des Zugangs zum Archivgut. Ebenso wie bei der Nutzung elektronischer Instrumente in der Büroarbeit schafft der Einsatz elektronischer Werkzeuge beim Zugang zu Archivgut so viele neue Möglichkeiten, daß eine neue Reflexion über Ziele und Methoden erforderlich wird.

Der Sinn des archivischen Findbuchs ist die freie und uneingeschränkte, assoziativ arbeitende Ermittlung bisher unbekannter Informationen. Bei der Findbucherstellung stehen zwei Instrumente für diesen Zweck bereit. Ihr Einsatz ermöglicht es, die inneren Zusammenhänge der Bestände trotz ihrer Verschiedenheit, immer wieder adäquat abzubilden. Das ist einerseits der Titel der Verzeichnungseinheit mit der Bezeichnung des Entstehungszwecks und seiner zeitlichen Identifizierung mit Hilfe der Laufzeitangabe und andererseits die Gliederung des ganzen Bestandes. Versuche, mit dem Titel eine Inhaltsbeschreibung zu geben oder ihn durch Deskriptoren oder Schlagworte zu ersetzen, sind gescheitert, weil sie notwendigerweise Auswertungsfragen antizipieren und privilegieren und anderen Fragen damit den Zugang verstellen. Der Titel bezeichnet das Ereignis, daß die Entstehung der Unterlagen verursacht hat. Alles andere muß der Interpretation der Benutzer überlassen bleiben. Die Gliederung hilft beim Verständnis der Titel einzelner Einheiten, indem sie die Kontexte erkennbar macht. Dadurch, daß sie vom Ganzen zu seinen Teilen fortschreitend eine Gesamtkompetenz vollständig in mehreren Ebenen unterteilt, wird der Umfang einzelner Gliederungsgruppen aus der Negation der übrigen Gruppen deutlich, ohne eine vollständige Aufzählung der möglicherweise dazugehörenden Fälle zu erfordern. Die beiden Instrumente, also die Titelformulierung und die Gliederung, sind zwar in ihrem Bedeutungsgehalt normiert, nicht jedoch in ihrer Anwendung. Deshalb können mit ihnen die Besonderheiten jedes Bestandes respektiert und präsentiert werden, ohne daß ihm ein vorgegebenes Schema übergestülpt werden muß.

Das Findbuch verhilft Benutzern auf zwei Wegen zu den gewünschten Unterlagen: Durch ermittelndes Schließen wird der Platz in der Struktur des Bestandes gefunden und die Beschreibung der Einheiten läßt Schlüsse auf die bei seiner Entstehung möglicherweise vorgenommen Ermittlungen und Erhebungen zu. So schaffen Findbücher Zugang zu Quellen neuen Wissens, das zuvor noch niemand bekannt war, auch nicht den erschließenden Archivaren. Sie kennen nur den Weg dazu, weil sie ihn freigelegt haben.

Archivische Findbücher sind bei der Archivbenutzung Instrumente eines Anregungen aufnehmenden, neue Ideen und Perspektiven entwickelnden Suchens und Denkens. Ihnen geht es nicht darum, eine bekannte Angabe durch normierte und deshalb überall gleichförmige Bezeichnung wieder auffindbar

zu machen. Dazu benötigt man andere Instrumente. Ihnen geht es um Hinweise auf Neuland. Sie sind Wegweiser für Entdeckungen. Deshalb nehmen sie die Inhalte möglicher Entdeckungen nicht vorweg. Bescheidenerweise benennen sie nur den formalen Rahmen und bezeichnen Gebiete für mögliche Nachforschungen. Gerade diese Funktionen können durch die Nutzung der nicht linearen Verknüpfungen codierter Texte besser dargestellt und leichter benutzbar gemacht werden als auf dem Papier. Sie können sie zudem etwa durch die Anbindung erläuternder Zusatztexte, Graphiken und Bilder an Stellen, an denen sie hilfreich sind, ohne beim schnellen Durchsehen zu stören, noch weiter unterstützen.

Online-Findbücher erweitern die Funktionalitäten bisheriger archivischer Erschließung. Sie geben neue Möglichkeiten der archivübergreifenden, gemeinsamen Präsentation von Beständen. So können virtuelle Verbünde zwischen Archiven hergestellt werden und mancher komplizierte Archivalienaustausch könnte sich bei einer Verknüpfung der Findmittel im Internet vielleicht erübrigen. Mit der Online-Präsentation ihrer Findmittel stellen sich die Archive gleichzeitig einem methodischen Wettbewerb, der zur ständigen Verbesserung der Erschließungsqualität genutzt werden kann. Das bedeutet eine neue Qualität des Zugangs zu Archiven. Da es ihr Ziel ist, Benutzern die Verwendung der inneren Strukturen von Archivbeständen zu erleichtern, statt sie vor ihnen zu verstecken, sollte ein Schwerpunkt bei der Weiterentwicklung der Online-Findbücher auf der Handhabung der Gliederung und ihrer übersichtlichen Darstellung auf verschiedenen Aggregationsebenen, vielleicht auch mit dem Einsatz neuer Visualisierungstechniken liegen.

Online-Findbücher und EAD kommen aus verschiedenen Zusammenhängen. Es ist aber heute wohl erkennbar, daß ihr Ansatz, obwohl sie sich in der Entwicklung nicht aufeinander bezogen, vergleichbar ist. Die Erfahrungen mit EAD werden sehr hilfreich für die Weiterentwicklung des Online-Findbuchs sein.

Die neuen Erschließungsformen liefern den Hinweis auf den zweiten Aspekt des aktuellen Paradigmenwechsels. Die Entwicklung der EDV-Anwendungen in den Archiven zeigt die zunehmende Hinwendung zu den Benutzern. Die Erschließung von Archivgut wird mit den neuen Instrumenten radikal nach außen orientiert und trotzdem, ja gerade dadurch werden die alten Prinzipien archivischer Arbeit neu begründet. Das Provenienzprinzip liefert die Begründung für die Struktur der Findmittel und für die navigierende Recherche. Es begründet eine Bewertungsmethodik, die der Beseitigung von Ballast und Redundanzen dient, und die eindeutige Präsentation und Öffnung der Bestände unterstützt. Eine neue Qualität des Zugangs und der inhaltlichen Auswertung wird möglich, gerade indem der archivischen Arbeit neue Hilfsmittel für eine durchgängige Orientierung an Formen und Strukturen bereitgestellt werden und sie sich zu den Inhalten neutral verhalten kann.

Archive als Anbieter von Informationspotentialen

Digitale Lesesäle, virtuelle Magazine, Online-Findbücher, alle diese Entwicklungen begründen offensichtlich keine völlige Neuorientierung der archivischen Arbeit. Allerdings tragen sie zur Klärung ihrer Zwecke und Funktionen bei und verschärfen die Frage nach dem Sinn des Berufes. Beide Seiten, sowohl das Verhältnis zur Verwaltung wie die breitere Öffnung zu den Benutzern bekräftigen die traditionellen Grundlagen der Archivwissenschaft. Sie bewirken keine radikale Änderung. Aber sie zeigen Archive als zunehmend aktivere Gestalter des öffentlichen Lebens. Dabei sind sie selbst an der inhaltlichen Ausgestaltung uninteressiert. Ihre Funktion als neutraler Anbieter von Potentialen, als Organisator von Zugängen, also sozusagen als Provider oder Enabler, begründet ihre Bedeutung.

Schon bisher haben die Bemühungen um die physische Erhaltung, um die Ausschaltung von Redundanzen bei der Bewertung, um die Präsentation und Bereitstellung den Zweck, die Realisierung der vollständigen und unverfälschten Aussagekraft des Archivguts möglich zu machen. Es geht also immer schon um die Informationspotentiale, die bisher allerdings nicht von den Gegenständen getrennt vorstellbar waren.[9]

Die Qualität der Informationspotentiale des Archivguts hängt direkt von der Erkennbarkeit der vollständigen Erscheinungsform, ihres ursprünglichen Aussehens und aller Veränderungen mit ihren zeitlichen Bezügen ab. Die analoge Reproduktion mit Hilfe von Mikrofilmaufnahmen zeigt das Prinzip, wie die vollständige Erhaltung der Erkennbarkeit auch bei Verlust der originalen Dinglichkeit möglich ist. Sie schafft nämlich ein Abbild, das sich selbst gleichzeitig mit dem abgebildeten Objekt als Reproduktion zu erkennen gibt. Ein Faksimile, eine Nachbildung oder eine Fälschung dagegen, bleiben selbst dann, wenn sie täuschend echt aussehen, unvollständig, weil sie sich nicht als Nachbildung zu erkennen geben. Das gleiche gilt für digitale Kopien. Von einer digitalen Aufzeichnung kann keine 100%ig identische Kopie hergestellt

---

9 Und trotzdem gab es Ansätze, diese Trennung zu denken. Adolf Brenneke konzipierte in seinen Lehrveranstaltungen der 30er Jahre das Provenienzprinzip als abstraktes Konzept, mit dessen Hilfe einem Bestand eine solche Form gegeben werden konnte, die am besten zeigte, wie er entstanden war, auch wenn er zuvor nie in dieser Form vorhanden gewesen war. Sein Ansatz, der Ähnlichkeiten mit der Systemtheorie des beginnenden 20ten Jahrhunderts in Wien mit ihren Wurzeln in der Biologie hat, war die Basis für die Unterscheidung zwischen Primärzwecken während der Entscheidungsfindung und Sekundärzwecken bei der Einsicht in eben diese Entscheidungen, wie sie Schellenberg formulierte. Beide Ansätze waren der sich auf Jenkinson berufenden Tradition entgegengesetzt, die den Perspektivenwechsel beim Übergang von der Verwaltung zum Archiv nicht akzeptierte. Vgl. *Cook*, wie Anm. 5, S. 19.

werden, wie Margaret Hedstrom dargestellt hat.[10] Deshalb wählt jede Migration Informationspotentiale aus und privilegiert sie gegenüber anderen, die vernichtet werden. Sie stellt eine Nachbildung her, die sich selbst aber nicht als Nachbildung zu erkennen gibt und ihre Entstehung nicht aufdecken kann. Das Prinzip der Mikrofilmaufnahme, die ihre Eigenschaft als Abbild gleichzeitig mit dem dargestellten Gegenstand präsentiert, schafft die Möglichkeit, die Ausgangssituation zu rekonstruieren. Auf diesen Umstand weist gerade ein etwa mitfotografierter Maßstab deutlich hin. Er verweist direkt nach außen und fordert dazu auf, die Abbildung nicht von ihrem Zweck zu isolieren, sondern ihre Funktion mit zu bedenken. An Hand der Proportionen läßt sich die ursprüngliche Größe trotz der verkleinerten Abbildung rekonstruieren, selbst wenn sie nicht nachgebildet wurde. Über die Rekonstruierbarkeit der Ausgangslage wird das Informationspotential vollständig erhalten, ohne daß bestimmte inhaltlich festgelegte Informationen privilegiert würden.[11]

Das Prinzip der Rekonstruierbarkeit erübrigt einen nicht begrenzbaren Aufwand für Nachbildungen des Originals, die doch immer unzulänglich sein werden. In der analogen Umgebung ist die dazu erforderliche Darstellung des Abbilds als Abbild selbstverständlich. Ihre materielle Beschaffenheit macht die Wahrnehmbarkeit der Unterschiede zwischen Original und Nachbildung zur Banalität. Diese Selbstverständlichkeit geht bei der Anwendung digitaler Aufzeichnungsformen verloren. Nicht das digitale Image oder die migrierte Kopie ist der modernere Ersatz für die Mikrofilmaufnahme, sondern das von Jeff Rothenberg entwickelte Konzept der Emulation statt Migration realisiert den Ansatz[12] der Rekonstrierbarkeit von Informationspotentialen für digitale Aufzeichnungen und formuliert damit ein grundsätzliches Element für eine archivische Strategie. Sowohl die Erhaltung des Originals wie die Ermöglichung der vollständigen ideellen Rekonstruktion seines ursprünglichen Zustands sind gleichwertige, nämlich funktional äquivalente Verfahren des Zugangs. Wenn die Ausgangssituation nicht erhaltbar ist, kann ihre Aussagekraft trotzdem durch ihre Rekonstruierbarkeit bewahrt werden. Von dieser Posi-

---

10 Margaret *Hedstrom:* Research Issues in Migration and Long-Term Preservation. In: Archives and Museums Informatics, vol. 11, No 3–4 (1997) S. 287–291.

11 Informationspotentiale liegen in der Evidenz begründet, die die Aufzeichnungen über ihre Verwendung geben. Diese offensichtliche Nutzung und die dadurch verursachte Veränderung, die ihrerseits eine Mitteilung für andere beteiligte Stellen war und von diesen in einem bestimmten, aus wiederum ihren Reaktionen erkennbaren Sinn verstanden wurde, kann nicht durch Metadaten ersetzt werden. Daten sind Angaben für bestimmte, genau bestimmte Informationsbedürfnisse, die zuvor antizipiert werden müssen, um im Vorgriff auf ihr Zutagetreten die Antworten auf Vorrat parat zu halten. Der Unterschied der Metadaten zur Evidenz liegt darin begründet, daß sie Antworten geben und nicht Interpretationsspielräume, also Antwortpotentiale eröffnen.

12 Jeff *Rothenberg:* Ensuring the longevity of digital documents. In: Scientific American (1995) p. 43ff.

tion aus können archivische Strategien entwickelt werden, die in einer elektronischen Verwaltungsumgebung nicht nur die offiziellen Dokumente, sondern auch die e-mails und andere Kommunikationsformen bei ihrer Vorbereitung berücksichtigen können und die den Benutzern der Archive die bestmögliche Offenheit ihrer Bestände anbieten können.

Diese Aufgabe ist allerdings etwas komplizierter, als es in der Formel von der Erhaltung des digitalen Erbes publikumswirksam verkürzt wird. Denn sie umfaßt die Entscheidung darüber, was digital und was in anderer Form erhalten werden muß, damit seine Informationspotentiale vollständig rekonstruierbar bleiben. Es sind flexible Strategien gefragt, die verschiedene Aufzeichnungsformen je nach dem verfolgten Zweck einsetzen.[13] Die Hauptaufgabe ist dann zu entscheiden. Das kann nur auf der Basis exakter Analysen geschehen. Regeln und Richtlinien sind dafür nicht ausreichend. Um mit solchen Situationen umgehen zu können, sind eine gute Fachausbildung und gründliche Kenntnisse erforderlich, damit eigene Strategien entwickelt werden können. Es müssen, ähnlich wie bisher schon in der Bewertung, mehr Werkzeuge vor allem für die Analyse entwickelt werden, die für konkrete Situationen bereit stehen. Sie benötigt Typisierungen von möglichen Erscheinungsformen und Beschreibungen von Konsequenzen möglicher Reaktionen, also die Aufarbeitung bisheriger Erfahrungen mit den verschiedenen Medien in einer wissenschaftlichen Theorie. Auf der Basis des übergreifenden Zieles, die Rekonstruierbarkeit der Ausgangssituation zu erhalten, kann sich ein neues professionelles Paradigma für digitale Archive entwickeln. Nebenbei zeigen einmal mehr die archivischen Prinzipien ihren unverzichtbaren Stellenwert für die Modernisierung des Berufs.

---

[13] Vgl. im gleichen Sinn Margaret *Hedstrom:* Building Records Keeping Systems: Archivists Are Not Alone on the Wild Frontier. In: Archivaria 44 (1997) S. 44–97, wo sie eine flexible Strategie fordert, die auch den Einsatz von Papier, wenn seine speziellen Funktionen gefragt sind, nicht ausschließt.

The Transition to Integration Incorporating the Digital Library into the University Research Library

Von Carol A. Mandel

Introduction

The documentation of knowledge comes in many forms, and for decades libraries have been incorporating new formats and new media into their collections. Microforms, audio recordings, video tapes and CD-ROMs have required specialized storage, equipment and cataloging techniques as well as new service methods and new access policies. However, none of these media have required radical change or reconceptualization of library practices or services. In contrast, online digital information appears to be a transformative technology with the potential to affect profoundly methods of research and, thus, the practice, and in some aspects, the theory of library service. Because of its dramatic potential, initial reactions to the concept of the „digital library" were to view it as a new, separate entity, a new world of information that would leave the traditional library behind as a dusty, unused archive. Experience of the last several years has instead demonstrated that for a long transitional period librarians will be delivering library information in an environment that must fully integrate both physical and virtual media. In American universities, as librarians have begun to incorporate online digital information into the university research library, they are currently experiencing a shorter term transition – the transition to this integrated state. This paper provides an overview of new trends and issues that are emerging from the early stages of the adoption of the digital library in the United States.

An operational view of the digital library

An element of the initial development has been to conceptualize an operational definition of the „digital library" that sets the stage for identifying and shaping needed services. The nature of digital information has so many implications for changes in patterns of information access, that it has taken many papers, meetings and discussions to separate out those aspects of the „digital library" most relevant to the mission and context of established research libraries. For example, from one perspective, often that of the general public, the

entire World Wide Web can be viewed as a rich and sprawling virtual library, in which hyperlinks and full text searches obviate the need for traditional modes of organizing information. From another perspective, that of the computer scientist, the term „digital library" is common jargon to denote a suite of sophisticated software programs for storing, managing and retrieving digital information. Often this perspective too assumes new technology as a replacement for library services, in this case focusing on function rather than content as the defining element of a library. An alternative, and equally limiting view, of the term „digital library" has been used by librarians focused on digitization projects who define a digital library only by content, as a specific collection of digital text, sound or images. All of these views are valid and meaningful in their appropriate context, but none provides the conceptual framework for a long range effort to ensure the continued delivery of digital information consistent with the mission of a research library.

Donald Waters, the Director of the Digital Library Federation, has developed the following working definition of „digital library" that provides a conceptual base for the program of the Federation and for the activities of its member libraries as they explore effective strategies and methods for integrating online digital resources into their operations:

- Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities *(http://www.clir.org/diglib/dldefinition.htm).*

This definition is, of course, essentially the definition of an operational research library – remove the word „digital" from the definition and it describes the mission of a research library. By basing the concept of a digital library in organizational function, the definition enables practicing librarians and system developers to focus on their role in digital library development. Their role is to create the technologies, techniques and resources that will enable operating libraries to bring digital resources to their user community.

This operational role will be further shaped and defined by the particular mission of a research library, by the nature of its collections and the requirements and expectations of its users. For example, a financial or corporate information center would be largely concerned with the provision of commercial databases of current information, and the focus of its digital library efforts would be on the effective licensing and delivery of these databases. An historical archive or special collection of physically recorded information, such as archival records, manuscripts, photographs, etc., would aim its digital library

goal at extending access to its collections by means of digitization and the creation of access databases (such as databases of finding aids or other specialized metadata). In contrast, data archives – such as repositories of social science survey data, scientific measurement data or electronic records of an organization – must concern themselves with developing techniques and technologies for the continuing availability of this information in the face of rapidly changing hardware, software and recording media. And the university research library must be concerned with all of these issues – and many more – in its effort to integrate digital resources into the teaching, learning and research activities of its community of students and faculty.

To serve that community, a university's digital library must be integrated fully with the information infrastructure of its parent institution. The digital library is not a freestanding entity, accessed only by specialized workstations on a dedicated network. Digital library resources must be readily and ubiquitously available at both institutional and personal workstations. This essentially means that digital library resources must be delivered through the World Wide Web, and aimed at a sufficiently low level of software client and related applications to constitute an appropriate „common denominator" for the great majority of users. To ensure that resources are available not only to those who can afford their own personal computer – and to ensure that resources are always readily available to users when they are in the library – a network of public access workstations and printers must be provided. All digital library resources must work effectively with this equipment. Access control, including authentication and authorization, must be both secure and easy, almost transparent, for authorized users to navigate. Digital resources should be provided in standard formats and structures that can be easily integrated into instructional applications and shared intra-institutionally among students and faculty. And, because a true digital library is a distributed resource, digital resources must be structured and delivered in a way that facilitate inter-institutional access. These requirements present challenges in the design, development and delivery of digital content beyond those faced by more narrowly construed digital libraries.

Another challenge for the operational university library is to integrate delivery of digital resources into its existing library services. Digital resources must be used side-by-side with traditional library materials, not in a separate library. The selection, cataloging, reference and preservation requirements of digital materials draw on the same professional expertise that is being applied to traditional materials. At the same time, the delivery of digital resources requires new services, new tasks, new skills, new jobs and new organizational approaches to providing library services. All aspects of library service are affected, including developing collections, organizing them for use, providing user services and preservation. For the past several years, many American university libraries have been experiencing the impact of these new demands and inventing the means to absorb them into routine operations.

Developing digital collections

The development of research library collections has always required multi-faceted approaches, ranging from straightforward purchases to complex exchange and gift arrangements. The digital library adds new methods and techniques to the domain of collection development, greatly expanding and changing the concept of building a library collection. Three new modes of collection development are licensing, digitizing and creating content – all very new roles for university research libraries.

For the great majority of research library users, the digital materials of most value in their educational and scholarly work are full text journals, reference works and indexing databases acquired from both commercial and non-profit publishers. Libraries have been acquiring published materials for over three hundred years, but the digital library environment creates a new constraint: these materials are typically licensed rather than purchased. The implications of such licensing are both mundane and profound. Most evident is simply the enormous workload entailed by the negotiation of license agreements. An acquisition transaction that was formerly handled by a routine, automatically generated purchase order now requires intensive professional oversight and time-consuming processing and management. Librarians must become well versed in the legal issues of licensing, and often require both legal and technical advice from experts elsewhere in the institution before a license is executed. Such digital resources are usually expensive, and require thoughtful reallocation of resource budgets. Advantageous pricing may be available through consortial arrangements, but such arrangements in turn require an added layer of investigation, management and negotiation. Eventually these workloads will be normalized and reduced by the growth of large third-party aggregators and service bureaus and by the wide acceptance of standard license agreements. But it is unlikely that licensing transactions will ever be as simple to manage as one-time purchases.

A more profound change in library operations stems from the very nature of licensing, as libraries will no longer physically, or virtually, own large segments of their collections, but will instead lease them on a renewing basis. The library thus becomes a collection service bureau for its users rather than a repository. While librarians are beginning to face this change operationally – i.e., by providing this service – they have not addressed the long term implications for future research and preservation. In this new world, a research library is no longer positioned to ensure enduring access to research information by caring for collections under its control. Instead librarians must meet this challenge in collaboration with publishers and computing professionals, working with the former to develop practices and agreements for archival access and with the latter to develop techniques to make such practices and agreements feasible. This collaborative perspective is beginning to emerge, but scaleable operational models are not yet implanted.

While published electronic journals and databases are the most heavily used resources in university digital libraries, for many librarians the defining content of the digital library is material that has been digitized from traditional formats. The attraction of the digital medium, particularly the ease with which it can be accessed and the flexibility of its presentation, has focused considerable attention on the methods and practices of digitization. As the initial flurry of experimental digitization projects has passed, attention is being paid not just to the question of how to digitize material, but why. Digitization has several selected, focused roles in a university library, and American libraries are in the early stages of exploring and understanding the effects of these roles. These roles include: preservation and access, specialized research support, instructional support, and support of digital library development.

Much publicity has been given to the benefits of digitally reproducing materials that are frequently used (or have the potential to be frequently used), but are too fragile or precious to be made available to more than a few carefully monitored researchers. The digital surrogate can be available on demand at workstations around the world while the original remains in a protected environment. The American Memory Project at the Library of Congress in Washington, DC has pioneered this concept, and has been widely imitated. However, digitization can be an expensive and even painstaking process; and management, presentation and long term maintenance of the digital files require an additional and continuing investment. In selecting materials for digital reformatting, it is important to analyze their likely future use relative to these investments. It is appealing to feature a prized special collection on the World Wide Web, but the effort should be clearly justifiable, whether by the public relations value alone or, more importantly, by the confirmed existence of a likely audience. While almost all material that has been added to library collections is retained, because of its significant research value, how many actual researchers have the interest and ability to make use of any particular body of specialized material? Experience has demonstrated that, even when a conversion project has external funding, the cost to the institution in overhead for technical and managerial staff is great. Given the many challenges of implementing digital libraries, the time and talent of these staff are a scare resource to be allocated with care. American research libraries have become increasingly concerned about careful selection for digitization, and model policy documents are available on the many library Websites. (The Research Libraries Group provides references on this issue at *http://lyra.rlg.org/preserv/joint/selection.html.)*

A number of special collections have been selected for digitization not only to make them more accessible to a known body of researchers, but because researchers have seen the potential for new modes of investigation and communication of knowledge. Such scholar-driven projects provide some early visions of the way digital libraries will be used in the not-too-distant future. One

example is the Advanced Papyrological Information System, APIS *(http://www.columbia.edu/cu/libraries/libraries/inside/projects/apis/index.html),* which is creating a federated digital library of papyri held in various collections in the United States and Europe, enabling comparative study and an overview of material previously impossible. The images of papyri will be accessed through a union catalog integrating metadata prepared at the various institutions and presented with links to textual databases and other electronic resources already available in the field. A somewhat similar project, the Digital Scriptorium *(http://sunsite.berkeley.edu/Scriptorium/)*, is creating a union catalog of metadata describing dated and datable Medieval and Renaissance manuscripts with links to selected images of manuscript pages. The focus here is not to create a digital library of complete manuscripts, but a research and reference tool of dated examples. These projects, which are requiring multi-institutional collaboration among scholars, librarians, imaging experts and computing professionals, point to both the new technologies and new relationships required for the creation of useful digital collections.

APIS and the Digital Scriptorium are serving as resources for instruction as well as for research, providing examples for study and teaching. For example, the Digital Scriptorium images are the „textbook" for a bi-coastal, interdisciplinary graduate seminar team-taught by professors from Columbia University and the University of California, Berkeley. Using the technologies of video-conferencing, digital libraries and distance learning, the seminar illustrates an emerging model for specialized graduate education. At a number of American universities, digital reformatting selections are most often driven by the need to use the materials in a particular course. And even outside of educational institutions, digitalization projects are increasingly geared toward use in instruction; for example, applicability to elementary school curriculum has become an explicit criterion for digitization of special collections at the Library of Congress. As is the case with research-driven projects, digitization projects that support instruction create new collaborations between librarians, computing professionals and teaching faculty.

At this early stage of digital library development, many digitization projects have been driven as much by the need to experiment as by the choice of content. For example, the Making of America phase one project, a partnership of the libraries of Cornell University and the University of Michigan *(http://moa.cit.cornell.edu/MOA/moa-main_page.html* and *http://www.umdl.umich.edu/moa/)*, attempted to select materials that would generate demand, but had to do so without prior experience of digital library use. The project, which initially created bit-mapped page images of 19th century journals and monographs relevant to American history and culture (many of these images have now been scanned to enable full text retrieval), has not actually provided a digital library which has been heavily used, but has generated a wealth of needed experience and expertise. Librarians have a great deal to learn about the organization, management and use of digital docu-

ments, and most of these lessons will need to be based on content created by digitization projects.

Digital online distribution is not only a stimulus to new ways of presenting historical documents, but to the publication of newly created content as well. Early experiences with online publishing indicate not only new formats and new uses for publications, but also the potential for new combinations of content and new business models for publishing them. For example, data from Columbia University's Online Books Evaluation Project [url] and experience gained from the University Press' new online publication, Columbia International Affairs Online,CIAO, *(http://www.ciaonet.org)*, point to new models of online scholarly products that combine a range of current awareness material, such as links to web sites and journal indexes, with new ideas represented in working papers and quickly produced conference proceedings, along with more traditional monographic content. Development of such products requires a team approach, drawing not only on traditional publishing expertise, but also on university computing resources and on librarians' first hand knowledge of online service delivery and use patterns. Stanford University Libraries' creation of the Highwire Press *(http://highwire.stanford.edu)* exemplifies librarians' new vision of the role they can play in scholarly publishing. Just as librarians have become engaged in new partnerships to convert material for research and instruction, they now have the opportunity to become involved in the publishing process as well. This is a rare moment for libraries to collaborate with university presses and scholarly societies, participating actively in shaping the digital products they will be collecting in the future.

Organizing and managing digital collections

As digital collections have, very recently, begun to grow at a rapid rate, libraries must quickly face the challenge of organizing and managing these collections for use. While web menus can provide convenient means for access to a relatively small, selected body of digital resources, the narrative menu approach does not scale for controlling and managing large, multi-disciplinary digital library collections. The operational demands of a university digital library require librarians to reexamine the traditional functions of providing intellectual access to resources and to managing collections. Librarians can draw on a base of established theory and practice, but new techniques and approaches are required.

Providing intellectual access to digital resources requires both an expansion of current cataloging practice and library-oriented development of World Wide Web retrieval techniques. Cataloging theory certainly can encompass the cataloging of digital „documents," but the actual practice requires specialized – and continually developing – expertise in describing new formats, features and relationships. Because the digital library brings together a wide va-

riety of resources, cataloging practice must be expanded to accommodate them. Alternative metadata models may be necessary for various materials; for example, both the APIS and Digital Scriptorium projects could not rely only on standard MARC-based, book-oriented cataloging and required data dictionaries suited to the specialized nature of the material. Metadata records in some form will continue to be a highly desirable means for gaining intellectual access to digital resources. In addition, these records will be supplemented by the online retrieval techniques natural to the World Wide Web, specifically full text searching and menu browsing. These techniques need considerable further development for research library application, and librarians must collaborate with Web developers and computing professionals to shape these services. The work of OCLC's Office of Research in collaborating with WC3 (World Wide Web Consortium) developers in designing the Resource Description Framework is a good example. (See, for example, paper by Eric Miller at *http://www.dlib.org/dlib/may98/miller/05miller.html/)* A welcome development for research libraries will be the implementation of XML and the ability to perform field specific searching on Web documents. Techniques for managing Web menus and for providing classified displays are in the pilot stages; these are essential developments for the true deployment of Web technology as a means for managing large multi-disciplinary research libraries of digital resources.

The level of workload and degree of complexity in managing digital collections are only now being glimpsed in an operational research library setting. Digital resources, whether they have been created locally and mounted on institutional servers or have been licensed for use from a remote server, are not static. To keep these resources readily accessible, a variety of digital collection management processes are required for inventory control, updating content, maintaining links, maintaining menus and presentations and providing security control such as authorization. As the number of resources increases, procedures based on checklists and menus will not scale as a means to managing collections. Today's library management software systems have been designed for oversight of physical books; the systems of tomorrow will need to handle digital collections.

Figure 1 displays a diagram of a process for digital collection management currently being piloted at Columbia University Libraries. (System designers are Stephen Davis, Director Library Systems, and David Millman, Manager of Research and Development for Academic Information Systems. Diagram created by S. Davis. Fuller descriptions can be found at *http://www.columbia.edu/cu/libraries/libraries/inside/projects/metadata/index.html.*). The process relies on a central relational database, the Master Metadata File, that can be used to track and monitor locally mounted digital files and remote databases. Inputs to the relational database are metadata from a variety of schemes and projects. Examples include: MARC records from the library's online catalog, modified MARC records from the APIS project, Digital Scriptorium re-

cords created in a Microsoft Access database, and „profile“ forms describing reference and indexing databases that have been created by librarians as HTML documents. The database will eventually be searchable directly. At this stage, it provides a central source for maintaining currency of information and generates updated Web pages for user access on the library's Web site. For example, standard outputs from the Master Metadata File include standardized HTML menus by author, artist and subject for various digital collections; customized HTML pages for browsing specialized collections; and HTML „connect“ screens for user access to reference and indexing databases. The Master Metadata File is described here not as the definitive system of the future, but as an illustration of the kinds of functions necessary in an active digital library environment. Librarians are only in the earliest stages of identifying the requirements for collection management in the operational digital library.

Assisting digital library users

All libraries should be designed with the goal of user self-sufficiency. In the traditional library, objectives toward this goal include such services as full cataloging, a „user-friendly“ online catalog, well-managed stacks, good signage, etc. Similarly a digital library facilitates independent use through well-organized and well-managed files and through well designed user interfaces, presentations and software applications. But in libraries that support research and instruction it is unrealistic – and counter to their mission – to underestimate the need for support and assistance to readers in finding and using both traditional and electronic resources. Casual Web browsing may be sufficient to identify useful resources for a quick (and potentially skewed) introduction to a topic, but research requires a more methodical and skilled approach to locating materials. Further, the educational mission of a university obligates its librarians to ensure that students gain some level of facility in accessing research information. No matter how well collections are managed and presented, readers will need help to take full advantage of them. Providing user support for access to digital research collections extends existing services and requires new ones. Programs of instruction and assistance must be expanded, and the entirely new function of usability testing now plays a growing role.

Programs of instruction in the use of library resources are well established in many university libraries, and vary widely depending on the needs and expectations of the university community. The introduction of digital resources adds a new layer, and possibly a new dimension, to instructional activities. In the current environment, the immediate challenge is to train both students and faculty in the discovery and use of resources and techniques that did not even exist a year or two ago. If the pace of new development in information technology continues, this will be an ongoing need. In the last several years, libraries in the U. S. have not only expanded instructional programs, but have had

to create new instructional facilities that enable computer projection and hands-on computer laboratories for training. These facilities in turn engender significant new workloads and expenses, as equipment must be continually maintained and upgraded. A particularly challenging problem has been to find appropriate training modes for faculty, who are disinclined to attend pre-scheduled group training sessions and require more individualized assistance for their specialized needs.

In fact, it is the necessity for more personal assistance for all library users, students and faculty alike, that presses for new models of user assistance in the digital library. There are many new demands. First, the digital library is open more hours than the traditional library – in fact, 24 hours per day, seven days per week. Second, its readers are not necessarily on site, and cannot present themselves at a reference desk whenever they need help. And finally, whether used on-site at public workstations or remotely from personal computers, the digital library requires a necessary interaction of content with hardware, software and network applications that complicate the process of finding and using resources. These issues require creative approaches to providing personalized, on-demand services in a feasible manner. New programs that have been implemented in the last several years in American libraries include a variety of online help and documentation services and the deployment of specialized personnel and trained student assistants located in residence halls and academic departments. For example, Columbia University Libraries and the University's Academic Information Systems jointly train a relatively small cadre of graduate students as Electronic Resources Assistants (ERAs) in various disciplines to provide assistance in the use of electronic information resources to graduate students and faculty in their academic departments. The program operates collaboratively with the schools and departments, and academic units provide office space so that the ERA is easily accessible to his/her colleagues. Each ERA works closely with a librarian who is a specialist in the discipline. The ERA's tasks range from providing individual consultations and small training sessions and developing documentation to helping academic departments to create Web sites and assisting individual users to connect to the campus network. Collaborative approaches to providing user assistance are essential in the digital library environment. The integration of digital resources with the electronic environment that delivers them often makes it difficult to untangle traditional library reference questions from calls for help in using dial-up software. American universities are implementing new help desk services that are supported by joint efforts from library, computing and telecommunications personnel.

Probably the most pervasive new workload generated from new digital resources is the continual need for testing and monitoring their usability. In many ways, these resources have as much in common with computer software applications as they do with the print resources that were their antecedents. These resources must be searched and displayed on computers, and, like any

suite of digital code, they may behave differently on different platforms. Even in the most closed and controlled environment, these resources would need to be monitored and maintained for new generations of operating systems, browsers and hardware – and a goal of the digital library is not to restrict distribution to a controlled environment. Libraries must ensure that all resources behave in expected and workable ways on university-supported public workstations, and test and monitor resource behavior on the (often many) platforms commonly used by their institution's students, faculty and staff. Compatibility with public workstations is often as, or more, problematic than with personal computers, since most commercial digital products are designed for common personal platforms, such as a current release of Microsoft Windows, rather than for university-based approaches, such as the UNIX environment. Support of printing on public workstations is a growing problem, and all of the various approaches – coin-operated „slave" printers, account-based network printing, free printing, etc.–currently employed in American university libraries are problematic in some ways. Some digital resources require additional implementation efforts, for example, delivery of specialized software for viewing or manipulating content. And licensing restrictions add requirements for security and authorization functions to restrict the use of products to a defined community. The overhead of these workloads has been enormous. In the long term, standardized solutions may emerge and the pace of change may either become slower (since we are currently in a period of experimentation rather than accepted practice) or become transparent (as software becomes more sophisticated in taking multiple platforms and generations into account). But this long term simplification is not guaranteed. Certainly in the near term transition to integration, the work of usability assurance is proving to be significant and is likely to grow.

Preservation of digital resources

The preservation challenges posed to libraries by digital formats are being increasingly well understood, thanks in large part to the work of the Commission on Preservation and Access of the Council on Library and Information Resources. The Commission's seminal report, Archiving Digital Resources, sponsored jointly with the Research Libraries Group, *(http://www.rlg.org/ArchTF/)* has been widely read and discussed by librarians, archivists and digital library technical experts. Its video production, Into the Future, *(http://www.clir.org/programs/otheractiv/intro.html*) is bringing an understanding of the nature and magnitude of the problem to a much wider audience. However, despite a growing recognition of the critical reality of the problem of digital resource preservation, this aspect of the digital library has not yet had an impact on research library operations. Why? Because librarians and computing professionals have not yet found an operational ap-

proach to addressing the problem. In the context of operational digital research libraries, two key issues shape the search for solutions. The first was described earlier: research libraries no longer own many of their most essential digital resources, but license them instead. Ensuring the future availability of these resources entails a collaborative process, engaging publishers both in issues related to the need for enduring access, and in techniques and processes for providing it. It is this latter requirement that constitutes the second key aspect of the problem: preservation of digital resources requires a set of ongoing, perpetual commitments, activities and processes, rather than a one-time, carefully controlled act of preservation. The activities and processes are necessarily complex, as they relate to the varying requirements for migrating different modes and formats of digital information into new generations of computer systems and software. Moreover, future migration requirements cannot be known at the present time, so that a commitment to preservation of digital resources requires a commitment to activities and processes not yet defined or attached to known costs. To the extent that an alternative to migration – emulation – has been proposed, it is an approach that at best will take many more years of research and adoption, leaving libraries with the current generation of digital resources to preserve by less sophisticated means.

In this early stage of understanding the challenge, libraries can begin by taking a few preliminary steps. These include:

1. identifying locally owned digital files of archival value for which the library has primary preservation responsibility;
2. ensuring that those files are in standard, non-proprietary formats;
3. to the extent possible, ensuring that the information content of the files is separable from any tools or applications used to manipulate it (i.e., store archival information in „flat files“);
4. maintaining content files in a format that can be handled by current software and hardware applications, and migrating those files to new formats as necessary;
5. implementing processes for the preservation of the physical media that contain the files, including back-up, use of high quality, durable media, appropriate environmental control and housing, periodic reading of the files and transfer to „fresh“ media as needed.

To the extent that an archival file does not separate content from software applications, individual approaches must be developed. All of these processes, both general and specific, necessitate the reliable assignment of staff and resources to the activity of digital preservation. These are activities that have not yet been routinely incorporated into operational research library preservation programs, even at libraries considered in the forefront of digital library development. Operational models cannot be put into place by the actions of preservation alone, and will in most organizations affect departments outside of

preservation – and, in many cases, outside of the library. Sharing information about working models and practices for digital resource preservation in university libraries will help effect the transition to integration in this area.

Organizational strategies for effecting the transition

The preceding overview of operational digital library functions has described new services and new tasks that have been integrated into the work of a number of American university libraries. In turn, the adoption of these activities has had a significant impact on the human resource requirements for these libraries and a more modest, but noticeable impact upon their organization. Given the extent to which digital services are being newly invented in these libraries, there has been a surprising degree of similarity and common experience.

The most obvious and immediate effect has been the urgent need for the new skills and expertise necessary to create and manage digital resources. Digitization programs, of course, demand new knowledge and capabilities related to the function of image capture and display, but the sets of skills necessary for digital library programs go far beyond the activities of scanning or digital photography. In fact, image capture functions are more easily contracted out to specialized vendors than those functions necessary to integrate the digitized files and other resources into library services. (Although in-house expertise is certainly necessary in order to specify and monitor vendor performance.) In addition to imaging expertise, necessary new skill sets include:

- expertise in providing intellectual access to digital resources, including specialized cataloging expertise and knowledge of metadata modeling and analysis;
- expertise in managing and presenting digital information, ranging from the ability to mark-up text in HTML and SGML to advanced knowledge of database management and programming skills for managing and presenting digital files;
- knowledge of intellectual property and contract law and related practices;
- advanced skills in techniques of instruction and presentation.

This expertise has largely been gained through staff development, both through training activities promoted by the institution itself and through individual librarians' initiative. Library and computing professionals have responded with dedication and energy to the professional development demands created by digital library functions, sponsoring a wealth of workshops, conferences, and training programs. Much of the knowledge needed has been gained through research and experimentation linked to follow-up activities to report out and share best practices. Professional organizations, including large

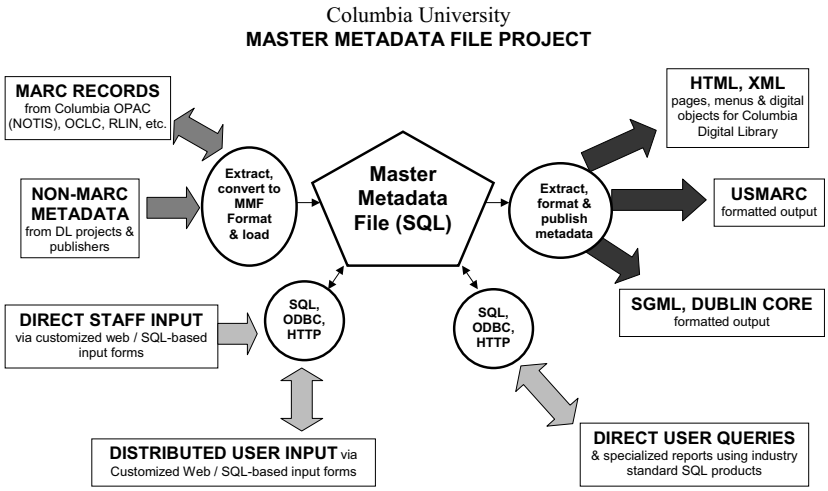
national associations such as the American Library Association and the Society of American Archivists, along with more specialized associations such as the Association of Research Libraries, the Coalition for Networked Information and the Digital Library Federation, have stepped up to the need for information sharing and training.

For any individual research library to obtain the mix of skills necessary to deliver digital library resources and services, senior management must focus planning and resources on developing existing staff and recruiting new ones. This has been an iterative process, taking advantage of and nurturing individual staff interest and talent, and then identifying and filling gaps. At this early stage of digital library implementation, a number of the most exciting programs in American libraries have stemmed from the talent and initiative of an individual librarian rather than from institutional foresight. As services mature and as these programs are adapted and refined by other institutions, digital library services are now more likely to be shaped by institutional planning.

As new services take shape, they generate new workloads, new job requirements and new position descriptions. Many new workloads and tasks are being absorbed into existing jobs: selectors are handling the assessment and usability testing of the electronic resources they acquire; reference librarians are playing a larger role in instruction; library systems staff are supporting access to digital resources as well as to the library's online catalog system. Additional staffing, often funded through reallocation from other tasks, is necessary to support exploding workloads in areas such as help desk support and equipment installation. New positions, whether supported by new funding or created through reallocation, are becoming necessary. Technical specialists positions are emerging in many libraries, particularly positions in imaging, database management, digital library programming specialties, and digital resources cataloging, as well as technical positions to develop complex Web sites, manage libraries of SGML-encoded documents, or oversee design and production for new digital publications. But many positions also reflect the need for leadership and coordination in order to develop new services and integrate them within a complex research library organization. Because much digital library work is still new and project-oriented, a commonly advertised position has a title such as „Digital Library Projects Coordinator." But positions that reflect a broader strategic commitment to digital library functions are also appearing in those libraries with the most aggressive programs, for example the position of Associate Director for Digital Library Initiatives at the University of Michigan, and the Chief Scientist position that leads digital library technical development at the University of California, Berkeley Libraries.

In almost all libraries that have implemented some level of ongoing digital library service, the organizational impact has been broader than tasks, functions and positions and has extended to organizational culture and modes of working. Digital library services are complex, and draw on both existing and new areas of expertise. For example, since digital resources are available in al-

most all fields of research, expertise in both subject and format skill are necessary to develop research library collections. The preservation of digital resources requires a combination of computing and preservation knowledge. During this transitional period – and perhaps ongoing – these combinations can be achieved only through teamwork and collaboration. Hallmarks of the transition organization include: interdepartmental teams and matrix management, new coordination roles, joint reporting relationships and new collaborations between libraries and other units of the university. For example, the new Imaging Coordinator position at Columbia University Libraries reports jointly to the Directors of Preservation and Systems; a new position of Archives Processing Coordinator reports jointly to the Directors of Bibliographic Control and Rare Books/Manuscripts, with a liaison relationship to Library Systems as well. Digital library development at Columbia is a joint effort of the Libraries and Academic Information Systems; online scholarly publishing projects result from the collaboration of these two units with the Columbia University Press; and a variety of specialized projects, such as the Digital Scriptorium, stem from collaborations with faculty. Similar examples can be found at other American universities that are aggressively working to effect the transition to integration of digital and traditional libraries. Digital library implementation goes beyond an internal transformation within the research library – in itself an exciting and challenging transformation – to changes in the role of the library within the university and the nature of its interaction with teaching, learning and research. The transition to integration includes a new model of integration of the research library with other parts of the university.



Der Einfluß digitaler Medien auf die bibliothekarischen Tätigkeiten

Von Hermann Leskien

Erläuterung

Einführung der Digitalen Bibliothek in großen Bibliotheken

Es ist wohl häufig zu konstatieren, daß griffige Schlagworte die Wirklichkeit nicht in hinreichender Differenzierung abbilden. Beim Gebrauch des Begriffs *Digitale Bibliothek* indessen wird ein besonders hohes Maß an Unklarheit ausgegossen – bis hin zu der Tatsache, daß manche derartig bezeichnete Software für alles andere taugt, nur nicht zur Verwaltung elektronischer Dokumente in Bibliotheken. Angesichts dieser Umstände ist es geboten, den Gegenstand des folgenden Beitrags knapp zu umreißen. Es soll ein Versuch sein abzuschätzen, welche Folgen sich ergeben, wenn man sich in einer großen klassischen Bibliothek darauf einläßt, elektronische Dokumente angemessen zu verwalten. Diese Überlegungen sind für den Autor keine theoretische Trockenübung, sondern fußen auf der Realität, der er sich in der eigenen Bibliothek gegenübersieht. Als eines der beiden Digitalisierungszentren der Deutschen Forschungsgemeinschaft ist die Bayerische Staatsbibliothek mehr noch als andere Einrichtungen in der Pflicht, die Auswirkungen der neuen Arbeitsmethoden auf das Bibliotheksmanagement und die bibliothekarischen Abläufe zu reflektieren.

Digitale Medien sind anders

Digitale Medien weisen in vielfacher Beziehung andere Eigenschaften auf als Drucke, gedeihen zumindest teilweise in einem anderen Umfeld und stellen Anforderungen, die wir in der Vergangenheit nicht gewohnt waren. Daraus folgt, daß mit den digitalen Dokumenten den in Bibliotheken vertretenen Informationsträgern keineswegs – wie dies noch bei den audiovisuellen Medien der Fall war – nur ein weiterer Typ hinzugefügt wird; vielmehr läßt sich, auch ohne ins einzelne zu gehen, feststellen, daß die damit geschaffene Lage um einiges komplexer ist. Nicht nur die Andersartigkeit der Medien, auch die gleichzeitig veränderte Welt der Informationstechnik, die schnellen Netze, die hohe Speicherdichte und die Diversifikation der Informationsmärkte stellen unsere überkommenen Vorstellungen radikal in Frage.

Arbeitsweisen

Die Kernfrage, die sich daraus ergibt, lautet: Werden die bisher durchgängig und mehr oder minder über die verschiedenen Medien einheitlich angewendeten Arbeitsweisen Bestand haben? Zumindest für die Verwaltung und Vermittlung elektronischer Dokumente sind Zweifel angebracht, aber möglicherweise bleibt der Prozeß nicht an der Grenze zwischen digital und nichtdigital stecken, sondern wirkt auf die Behandlung der Druckschriften zurück.

Eigenschaften digitaler Medien

In diesem Absatz sollen lediglich einige wichtige Eigenschaften ins Gedächtnis gerufen werden, die digitale Medien von Printmedien unterscheiden und letztlich *schuld* daran sind, daß sich möglicherweise die Geschäftsabläufe in Bibliotheken ändern.

Beliebige Verfügbarkeit

Zunächst ist die beliebige Verfügbarkeit zu nennen. Theoretisch – und bei guter Infrastruktur immer häufiger auch praktisch – macht es keinen Unterschied, ob von einer deutschen Bibliothek auf einen kalifornischen Server oder auf einen Server in der eigenen Bibliothek zugegriffen wird. Die ubiquitäre Verfügbarkeit eröffnet von Bibliothekaren bislang ungedachte Alternativen der Informationsbeschaffung, die es unter Aspekten der Wirtschaftlichkeit gegeneinander abzuwägen gilt. Die relative Schnelligkeit der Datenübermittlung kann gelegentlich – insbesondere bei Unsicherheit in der Einschätzung des Bedarfs – sogar die Sinnhaftigkeit eines Erwerbs überhaupt in Frage stellen.

Dynamik und Flüchtigkeit

Zu den Schattenseiten digitaler Medien gehört – zumindest im Fall der Netzpublikationen – die Instabilität der Dokumente. Sie sind nicht nur in sich fluid, wobei die früheren Fassungen keine Spuren hinterlassen, sondern sie haben oftmals insgesamt nur eine kurze Lebensdauer. Die Lebenserwartung einer Web-Seite erreicht im Schnitt noch nicht einmal zwei Monate. Da sie gemäß einer Statistik vom Januar 1997 durchschnittlich 15 Links zu anderen Seiten oder Objekten aufweist, zeigt sich auch aus dieser Warte die Brüchigkeit der komplexen digitalen Geflechte.[1]

---

[1] Brewster *Kahle* und Peter *Lyman:* Archiving Digital Cultural Artifacts. Organizing an Agenda for Action. 1998. – URL: *http://www.archive.org/us/brewster/cultural-artifacts.html*

Heterogenität der Informationsträger und Systeme

Trotz aller Normierungen und trotz aller Quasi-Standards verzeichnen wir in der elektronischen Welt ein hohes Maß an Heterogenität. Das trifft schon auf die Informationsträger selbst zu und setzt sich im laufenden Wechsel der Betriebssystem-Versionen fort. Hieraus resultieren Probleme bei der Sicherung langfristiger Verfügbarkeit. Aber es werden auch Kompatibilitätsprobleme schon am Tag des Erwerbs aufgeworfen. Die Sehnsucht nach einer Welt, die auf diesem Sektor mehr Uniformität garantiert, bleibt aber wohl ein auf Dauer unerfüllter Wunsch.

Tiefe, aber ungeregelte Erschließung

Wenden wir uns der Charakterisierung der Erschließung zu, so gilt es, gleichermaßen an Positives wie an Negatives zu erinnern. Ein elektronischer Text ist heute bis auf Wort- und Zeichenebene erschlossen und somit mühelos recherchierbar. Die dokument- und dateiüberschreitende Suche gesellt sich als besonderer Vorzug hinzu. Auch die Möglichkeit des raschen Überblicks und des beliebigen Wanderns zwischen Inhaltsverzeichnis und Text stellen gängigen Standard dar. Auf der Seite der Passiva finden sich vor allem die Folgen einer im wahrsten Sinn des Wortes 'regel-losen' Erschließung. Die von uns Bibliothekaren so nachhaltig verfolgte Politik der Normierung ist im Internet ein Fremdwort – wenngleich die Sehnsucht danach aufgrund der unzureichenden Suchergebnisse immer mehr wächst.

Mehrwert als neuer Vorzug

So wird man zusammenfassend sagen können, daß sich elektronische Dokumente mittel- bis langfristig dann durchsetzen und Druckschriften sowie andere Informationstypen verdrängen werden, wenn sie deutlichen Mehrwert erkennen lassen. Und diesen medienbedingten Mehrwert gibt es in der Tat. Einmal ist es der weltweite, liberalisierte Zugang, ein anderes Mal die tiefere Erschließung, ein drittes Mal die leichte Verarbeitbarkeit, ein viertes Mal sind es die niedrigeren Gestehungskosten, die digitale Formen den konventionellen überlegen machen.

Begleitphänomene

Den immanenten Eigenschaften seien einige Phänomene beigesellt, die wir im Umfeld der digitalen Medien verstärkt finden.

Medienpluralität

An erster Stelle ist die Medienpluralität zu nennen. Neben die bisher bekannten Typen treten die verschiedenen Formen der elektronischen Medien, die immer neue Gestalt bis hin zur gewohnten Papierausgabe annehmen können. Die Möglichkeit, vom Original abweichende Sekundärformen zu schaffen, trägt zur weiteren Vielfalt bei. Wir werden damit langfristig leben müssen; denn die euphorische Vision, die gesamte Welt elektronifizieren zu können, hat sich rasch als einfältiger Traum erwiesen. Medienbrüche effizient zu managen, entwickelt sich hingegen zur Daueraufgabe.

Anbieterpluralität

Auf Anbieterseite gibt es eine breite Palette von Akteuren, so daß man sagen kann: Die Zeit des Bibliotheksmonopols – wenn es denn so etwas jemals wirklich gab – ist endgültig vorbei. Mit Informationen handeln kann jeder; und viele tun es auch. Wettbewerb ist angesagt. In diesem Zusammenhang gewinnen Zugang, Aufbereitung, Präsentation und Vernetzung zunehmend an Bedeutung. Innovation, Mehrwert und Wettbewerbsvorteil liegen nicht selten gerade in diesen Bereichen und nicht beim bloßen Angebot der Information selbst.

Größere Kundennähe

Die leichte Verarbeitbarkeit und die Eignung zur schnellen Kompilation sowie die Vielfalt der Präsentationsmöglichkeiten führen im Verein mit den reduzierten Herstellungskosten in der Tendenz dazu, der Vielfalt auch in dieser Hinsicht aufzuhelfen. Die Entwicklungen fördern die Ausrichtung der Angebote auf spezielle Zielgruppen bis hin zur Individualität. Der Absatzmarkt wird dementsprechend kleinteiliger. Im Bereich der Wissenschaft bedeutet dies, daß eine größere Nähe zu den jeweiligen Wissenschaftsdisziplinen möglich wird und daß fachlich differenzierte Nachfrage nun zunehmend durch ein jeweils angemessenes Angebot befriedigt werden kann. Das Informationsverhalten der einzelnen Wissenschaften wird damit wohl immer weiter auseinanderlaufen – zum Leidwesen universal ausgerichteter Bibliotheken, die traditionell alle und alles über einen Kamm scheren wollen.

Ressourcen-Management

Aus den vielen Aspekten, die einer näheren Betrachtung wert wären, seien zwei Komplexe von zentraler Bedeutung herausgegriffen: Ressourcenmanagement und Erschließungsmanagement. Der Grund für diese Auswahl ist, daß

gerade auf diesen beiden Gebieten alte und bewährte Traditionen in Frage und zur Disposition gestellt werden. Die Fähigkeit des Wandels wird auch hier über das künftige Schicksal der Bibliotheken entscheiden.

Bestandsentwicklung

Ressourcenmanagement ist zunächst nichts anderes als der traditionelle Bestandsaufbau. Der Erwerb für die eigenen Regale war bisher die einzige Möglichkeit, ein Werk zu gegebener Zeit – entweder sofort oder für spätere Generationen – verläßlich verfügbar zu haben. Dem physischen Bestandsaufbau und der Bestandsentwicklung wird auch künftig ein hoher Stellenwert beizumessen sein. Das gilt gleichermaßen für alle Informationsträger und Publikationsformen. Denn große Bibliotheken werden angesichts der Massennachfrage gut daran tun, ihre Leistung wesentlich auf eigene Bestände zu gründen. (Anders sieht es bei Firmenbibliotheken aus, die mit wenigen Nachschlagewerken und guten Kontakten zu Content-Providern durchaus erfolgreich operieren können.)

Kooperativer Bestandsaufbau

Zwar noch recht physisch, wenngleich schon ein wenig virtueller, geht es beim kooperativen Bestandsaufbau zu. Im Rahmen eines Leistungsverbundes darauf zu zählen, daß im Bedarfsfall auf eine Partnerbibliothek zurückgegriffen werden kann, setzt wirtschaftliches Denken, allseitige Disziplin und hohe Verläßlichkeit voraus. Dies bezieht sich nicht nur auf die gegenwärtige Leistungsfähigkeit des jeweiligen Partners, sondern vor allem auch auf die Zuverlässigkeit langfristiger bis dauerhafter Vorhaltung. Kooperativer Bestandsaufbau schließt andererseits die Bereitschaft zum Verzicht und großräumiges Denken ein und wird in dem Maß gefördert, in dem wirksame Nachweissysteme – im Verein mit effizienten Dokumentlieferdiensten – an Wirksamkeit gewinnen.

Selektiver Erwerb im Bedarfsfall

Eine Variante oder zumindest eine Folge kooperativer Erwerbung stellt der Kauf im Bedarfsfall dar.[2] Nur selektiv angewandt, ist er wirtschaftlich. Den Begriff Erwerb sollte man jedoch in diesem Zusammenhang nicht allzu wört-

---

[2] Besonders bekannt geworden ist das Beispiel der Louisiana State University: Jane P. *Kleiner* und Charles *A. Hamaker:* Libraries 2000 – Transforming Libraries Using Document Delivery, Needs Assessment and Networked Resources. In: College & Research Libraries (1997) S. 355–374.

lich nehmen, denn eigentlich handelt es sich um die Vermittlung nachgefragter Dokumente ohne vorsorgliche Vorhaltung. Der kopierte Aufsatz etwa geht an den Besteller und verbleibt nicht in der Bibliothek. Allerdings ist es eine Form, die bei geschickter Anwendung unterm Strich deutlich Geld sparen hilft. In jedem Fall eröffneten sich derartige Möglichkeiten erst, seit es Online-Kataloge, Direktbestellungen und elektronische Lieferung gibt.

Lizenzerwerb

Neu im Spektrum des alt-ehrwürdigen Erwerbungsbibliothekars nimmt sich das Geschäft der Lizenzverhandlungen und des Lizenzerwerbs aus. Im Grund genommen handelt es sich um eine im Vergleich zum Bestandsaufbau geradezu konträre Art der Geldausgabe. Denn im einen Fall wird Geld für den Erwerb physischen Eigentums eingesetzt, im anderen gerade diese Form des Erwerbs eingeschränkt, um Zugangsrechte zu finanzieren, die sich am Ende möglicherweise überhaupt gar nicht rentieren. Nichtsdestoweniger handelt es sich um eine alternative Form der Sicherung des Zugangs zu Informationen und gehört damit genuin in das Aufgabenfeld der Abteilung.

De-Acquisition (Kassation) und Langzeitverfügbarkeit

Zum Ressourcenmanagement gehört auch der Aspekt der langfristigen Profilpflege des Bestandes, der zugleich den Grundstock für die Leistungsfähigkeit in der Zukunft bildet. Dies ist freilich nicht neu. Die Sorge um die Bestandserhaltung im rein physischen Sinn, der Versuch, im Bestand Lücken zu füllen, die in der Vergangenheit aus welchen Gründen auch immer gelassen wurden, oder das Ausscheiden von nicht archivwürdigem Bibliotheksgut sind klassische Aufgaben. Die digitalen Medien bringen nun – bedingt durch die Flüchtigkeit magnetischer Speicherung und durch die kurzen Zykluszeiten von Hard- und Software – nicht nur eine Intensivierung des Problems. Da kooperativer Erwerb und Lizenzerwerb den Verzicht auf die eigene Hausmacht einschließen, kommt darüber hinaus ein durchaus neuer Zug ins Spiel: die Vorhaltung und laufende Anpassung des Leistungsnetzes, auf das sich eine Bibliothek bei der primären Erwerbungsentscheidung gestützt hat.

Managementaufgabe Alternativenbewertung

Ziehen wir ein Resümee aus dieser knappen Analyse, so zeigt sich, daß die Aufgaben des klassischen Erwerbungsfachmanns vielfältiger geworden sind. Neben die Selektion der geeigneten Titel ist die Selektion der Informati-

onsträger, der Rechte und der Verfahren getreten. Diese Aspekte sind unter ökonomischen Gesichtspunkten und aus dem Blickwinkel aktueller und langfristiger Versorgung zu bewerten und nehmen wesentlichen Einfluß auf die Entscheidungen. Das Ziel ist eine Optimierung des Mitteleinsatzes; in Zeiten immer knapper werdender Ressourcen und immer stärker steigender Preise handelt es sich dabei um eine Managementaufgabe ersten Ranges. Ressourcenmanagement ist somit etwas anderes und weit mehr als schlichter Bücherkauf.

Erschließungsmanagement

Sinnvoll an Informationen heranzuführen, ist als vielschichtige Aufgabe anzusehen, die letztlich nur unter Beteiligung der Benutzer – d.h. der Fachwissenschaft oder der jeweiligen Zielgruppe – erfolgreich konzipiert werden kann. Wie ein Inhaltsverzeichnis zu präsentieren ist, wie Browsing- und Retrievalmöglichkeiten zum Volltext oder zum Image führen, das sind Grundlinien komplexer Erschließungskonzepte. Wer mit digitalen Medien umgeht, importiert neue Sichtweisen in das Erschließungssystem seiner Einrichtung. Traditionell ist das Retrieval von Zeitschriftenaufsätzen zwar in Bibliotheken möglich, jedoch nicht aufgrund von Erschließungsaktivitäten der Bibliotheken selbst. Unter Einbeziehung von Dokumentationsdiensten (Indexing Service) und des Dokuments selbst bietet sich verallgemeinert in der Regel folgende Datenverteilung:

Element	Bibl.-Katalog	Indexing Service	Original-Dokument	Digitale Version
Übergeordneter Titel	+	–	+	+
Bandaufführung	+	–	+	+
Inhaltsverzeichnis	–	–	+	+
Aufsatztitel	–	+	+	+
Systematisches Retrieval	–	+	–	–
Verbales Retrieval	–	+	–	–
Abstract	–	+	+	+
Volltext (Image)	–	–	+	–
Volltext (Zeichen)	–	–	+	+

Diese Übersicht macht klar, daß die herkömmliche Erschließung nur durch die Kombination verschiedener Instrumente (Katalog und Bibliographie/Datenbank) zum Standort und schließlich zum Dokument selbst führte. Erst die

Einsichtnahme in das Dokument führte zur Vollinformation. Bei digitalen Medien ist letztere im vollem Umfang möglich. Links führen ggf. von einem Instrument zum anderen, und die Einsichtnahme in ein physisches Original erübrigt sich im Regelfall.

Die elektronischen Medien halten also der klassischen Erschließung den Spiegel vor. Sie werden deshalb auch konzeptionell Rückwirkungen auf die Tradition haben, es sei denn, wir erliegen der Versuchung, die neuen Möglichkeiten in die alten Betten zu zwängen.

(Integration von) Formal- und Sacherschließung

Wenn wir von integrierten Geschäftsgängen sprechen, meinen wir gemeinhin die koordinierte Zusammenführung von Erwerbungs- und Katalogdaten in schlanken Geschäftsabläufen. An einigen Orten erfolgt darüber hinaus die konsequente Integration von Formal- und Sacherschließung. Zu nennen ist hier beispielsweise die Universitäts- und Landesbibliothek Düsseldorf. Hier wird – gestützt auf die Nutzung von Fremdleistungen – die Nutzung der Stichwortrecherche für die verbale Sacherschließung realisiert.[3] Die im übrigen eher schleppende Entwicklung in diese Richtung ist nicht nur darauf zurückzuführen, daß die Formalerschließung eine Domäne des gehobenen Dienstes und die Sacherschließung eine des höheren Dienstes ist; wenngleich dies tarifrechtlich durchaus eine heikle Problematik sein kann. Es liegt auch daran, daß wir – zumindest in Deutschland – keine Einheitlichkeit in der Sacherschließung haben. Sacherschließung meint einmal die bloße Möglichkeit der Stichwortrecherche, ein anderes Mal Schlagworterschließung (nach RSWK), ein drittes Mal systematische Erschließung (nach lokal oder regional angewandten Systematiken). Die Folge ist, daß – anders als bei der Titelaufnahme – das Ausmaß nutzbarer Fremdleistungen als gering eingeschätzt werden muß.

Es wäre durchaus möglich und ist gelegentlich auch schon angedacht, aktuelle Leistungen anderer Bibliotheken oder Informationsdienste auf dem Gebiet der Sacherschließung sowie den Inhalt historischer Kataloge zu nutzen. Aus wirtschaftlichen Gründen bedingt dies allerdings den Verzicht auf Einheitlichkeit in der Erschließungsmethodik und -tiefe. Die Folge könnte sein, daß in ein und derselben Bibliothek – sammlungs- oder fachgebietsbezogen – unterschiedliche Erschließungskonzepte verwirklicht werden. Sie würden zu einer Diversifikation der Modelle führen, andererseits aber u.a. auch eine Integration von Formal- und Sacherschließung ermöglichen und darüber hinaus

---

3 Zuletzt: Klaus *Lepsky:* Sacherschließung ohne RSWK? Neue Praxis an der Universitäts- und Landesbibliothek Düsseldorf. In: ProLibris (1998) S. 112–114.

sachnäheren Zugang bieten. Dieser Gewinn bezieht sich keineswegs nur auf einen effektiven Geschäftsgang, sondern – wie vorhin schon angedeutet – auf Erschließungskonzepte, die jeweils optimal an den Bedarf der betroffenen Klientel angepaßt sind.

Rolle eines sauberen Katalogs (in heterogenem Umfeld)

Die Einheitlichkeit von Katalogen ist ein hohes Gut, ohne Zweifel. Einheitlichkeit im Sinn von durchgängig einheitlicher Konzeption und verläßlicher Anwendung der gewählten Prinzipien in allen Fällen. Um es praktisch zu sagen: Ein Katalog ist dann ein guter Katalog, wenn der Informationssuchende präzise weiß, was die Antwort auf seine Frage wert ist. Das hohe Gut einheitlicher Kataloge muß aber heute insofern in seiner Bedeutung hinterfragt werden, als sich das Umfeld geändert hat – nicht erst seit dem Siegeszug des Internet, zugegeben, aber seitdem in unübersehbarer Weise.

Ein *sauberer* Katalog im eben beschriebenen Sinn ist auch heute noch der Königsweg, wenn die darin nachgewiesenen Bestände den Informationsbedarf der betroffenen Klientel umfassend befriedigen. Nur und gerade unter dieser Kondition kommen die Vorzüge der Prinzipien, die der Erschließung zugrunde liegen, voll zum Tragen.

Zwei konträre Beispiele mögen die Spannweite angemessener Lösungen illustrieren. Angenommen, es gibt einen Bereich, in dem die digitalen Medien gegenwärtig (noch) eine untergeordnete Rolle spielen; in der Geschichtswissenschaft etwa scheint dies der Fall zu sein. Hier macht es durchaus Sinn, auch digitale Dokumente in das traditionelle Erschließungssystem einzubeziehen. Der Katalog einer großen Bibliothek könnte damit für informationssuchende Historiker *der Nabel der Welt* sein. Betrachten wir hingegen die Realität im Bereich Medizin, so stellt sich die Sachlage ganz anders dar. Rund 90% des Informationsbedarfs zielen auf Aufsatzliteratur. Die diesbezüglichen Defizite in Katalogen großer Bibliotheken, die diesen Literaturtyp überhaupt nicht nachweisen, sind allgemein bekannt. Unter diesen Umständen in die Aufrechterhaltung der Einheitlichkeit der Sacherschließung personelle Ressourcen zu investieren, ist schlichtweg Vergeudung von Arbeitskraft; zumal wenn man bedenkt, daß das bei Medizinern gebräuchliche Vokabular im MeSH (Medical Subject Headings) mit dem deutschen Regelwerk RSWK nichts, aber auch gar nichts, gemein hat.

Die Reihe ließe sich fortsetzen. In jedem Fall scheint nur noch im Bereich der Geistes- und Sozialwissenschaften die Sinnhaftigkeit eines konsequent und einheitlich geführten Katalogs gegeben zu sein. Unsere gegenwärtige Aufgabe besteht darin, den Informationsbedarf konsequent aus Nutzersicht zu analysieren. Da diese Sicht in der digitalen Welt eine weit größere Rolle spielt als in der analogen, werden aus gründlichen Analysen Folgerungen zu ziehen sein. Haben wir im Fach A textdominierte Monostrukturen, sollten

wir – zumindest in großen Bibliotheken – die Traditionen weiterpflegen. Lebt eine Sparte B hingegen in einem Umfeld, das von Bibliotheken nicht zentral beherrscht wird und wo digitale Informationen einen besonders hohen Stellenwert haben, erscheint es unsinnig, an den alten Prinzipien festzuhalten. Wer mit kleinen Marktanteilen in einem heterogenen Umfeld agiert und danach strebt, Qualitätsmaßstäbe aufrecht zu erhalten, die von den Kunden nicht honoriert werden, legt den Grundstein für seinen Bankrott. So das kleine Einmaleins der Märkte.

Bibliothekare sollten nie vergessen, daß die wissenschaftlichen Benutzer ohnehin mehrere Bibliotheken und eine Vielzahl von Erschließungsinstrumenten zu nutzen gewohnt sind. Man darf ruhig auf die Intelligenz, die vor dem Bildschirm sitzt, vertrauen und sollte nicht der Versuchung erliegen, alles Heil von der Funktionalität von Systemen zu erwarten und daher ausschließlich in diese zu investieren. Wissenschaftler können mit Heterogenität sehr wohl umgehen, nur wir Bibliothekare gelegentlich nicht.

Vielfalt der Dokumente (und die Erschließung aller Dokumenttypen)

Zum Umdenken sind wir auch deshalb gezwungen, weil von uns künftig mehr erwartet wird, als unsere Erschließungsarbeit auf die Katalogisierung von Büchern und von Zeitschriftentiteln zu konzentrieren. Wir haben zwar schöne einheitliche Kataloge, erschließen darin aber Informationen bisweilen in einer völlig unangemessenen Form – so bei Aufsatzliteratur – oder gar nicht – so bei Faktendaten, aber auch Bildern, Objekten, Handschriften, Nachlässen oder Archivalien ganz allgemein. Folglich haben wir die entsprechenden Kundenkreise verloren oder gar nicht erst gewinnen können. Erfreulicherweise gibt es auf diesem Gebiet Fortschritte. Ein besonders auffälliges Beispiel liefert die Research Libraries Group, die mit den Projekten AMICO und AMC museales bzw. archivalisches Gut in die allgemeinen Erschließungsinstrumente einbezieht und integriert.

Kooperative und heterogene Erschließung

Diese Aussagen leiten nahtlos auf ein neues Feld über. Bei einer verteilten, kooperativen Erschließung bieten sich zwei grundsätzlich unterschiedliche Ausformungen, die auf den Prinzipien der Prä- und der Postkoordination beruhen. Im ersten Fall handelt es sich um eine mehr oder minder sorgsam geplante gemeinsame Anstrengung mit verteilten Rollen. Durch die Zusammenführung der auf diese Weise erstellten Erschließungsleistungen zu einem einheitlichen Ganzen entstehen auf kostengünstigem Weg leistungsfähige Instrumente. Die einschlägigen Aktivitäten auf dem Gebiet der Sacherschließung in Deutschland und anderswo haben die Nützlichkeit in den letzten Jahrzehnten eindrucksvoll unter Beweis gestellt.

Wenig verbreitet ist hingegen der Ansatz der Postkoordination von bereits vorliegenden Erschließungsleistungen oder von aktuellen Erschließungsaktivitäten, die sich – aus welchen Gründen auch immer – nicht zum Zeitpunkt der Entstehung koordinieren lassen. Die hohen Kosten einer Konversion nach bewährtem Muster der Einarbeitung standen und stehen einer Realisierung hindernd im Weg. Bei Aufgabe des Prinzips der Homogenität eröffnen sich jedoch neue Chancen. Die Frage lautet nämlich unter dieser Prämisse, ob heterogene Erschließungssysteme besser sind als gar keine oder solche, die nur in Karteiform vorliegen. Zu denken ist in diesem Zusammenhang an alte Sachkataloge und Aufstellungssysteme ebenso wie an verbale und systematische Erschließungsleistungen in anderen Bibliotheken des In- und Auslands. Erschließungen auf Aufsatz- und Kapitelebene können ebenso berücksichtigt werden wie Volltext-Recherchen bei digital vorliegenden Texten.

Aus dem Gesagten geht hervor, daß Kooperation und Nutzung historischer Leistungen keine Felder darstellen, auf denen Rationalisierungsgewinne entstehen, die zu Personaleinsparungen führen. Im Gegenteil muß einiger zusätzlicher Aufwand investiert werden, um zumindest minimale Brücken zwischen den heterogenen Systemen zu schlagen. Bei geeigneter Sachlage ist der Gewinn aber weit größer als der Aufwand, und insofern handelt es sich um höchst wirtschaftliche Projekte.

Besitzunabhängige Erschließung

Zu erwägen ist in dem einen oder anderen Fall auch der Import bzw. das Angebot von Erschließungsdaten, die Dokumente nachweisen, auch wenn sie sich nicht in der eigenen Bibliothek befinden. Natürlich haben wir das schon immer durch Vorhaltung von Bibliographien und von Katalogen anderer Sammlungen praktiziert. Heute geht es – in Fortsetzung dieser ererbten Aufgabe mit aktuellen Mitteln – um die sinnvolle Integration dieser Instrumente in das Informationsangebot einer Bibliothek. Gemeint ist damit eine Integration im engeren Sinn, d.h. eine Verknüpfung und Verflechtung mit eigenen Erschließungsdaten.

Natürlich stellt sich – ist das Prinzip erst einmal akzeptiert – sogleich die Frage, in welchen Fällen eine Erschließung über den eigenen Bestand hinaus tunlich ist und in welchen nicht. Da derartige Entscheidungen, wie Erwerbungsentscheidungen, nicht aus bloßem Gefühl heraus zu treffen sind und zumindest die Summe der Einzelfälle eine Linie erkennen lassen muß, kommt keine Institution umhin, ein Angebotsprofil für erschlossene Dokumente zu entwickeln. Dabei ist natürlich ein Konnex zu den Erwerbungs- und Vermittlungsentscheidungen gegeben. Zunächst einmal wird man sinnvollerweise Dokumente erschließen, die man eigentlich – wäre der Etat ausreichend – gern selbst erworben hätte. Es liegt darüber hinaus nahe, ein Fachprofil zu entwickeln. Eine Bibliothek nach diesem Zuschnitt ähnelt einem Fachgeschäft,

das nicht jede Ware auf Lager hat, aber neben dem Angebot im Hause gute Lieferbeziehungen nach draußen unterhält.

Diesen Prinzipien folgt in gewisser Weise die Deutsche Forschungsgemeinschaft mit dem pilothaften Aufbau von Servern an einigen Bibliotheken. Die Fachserver lehnen sich an das Sondersammelgebietsprogramm an, sind aber keineswegs an den Eigentumserwerb von Dokumenten gebunden. Es ist leicht vorherzusagen, daß sich die Server im Service fächerspezifisch auseinander entwickeln werden. Ein Ziel müssen und werden sie jedoch konsequent verfolgen: das jeweilige Benutzerbedürfnis zu befriedigen, indem Informationsbedarf und Dokumentlieferung gleichermaßen bedient werden.

Automatische Verfahren

Angesichts der heterogenen Welt, der wir uns gegenübersehen, kann man ein gewisses Verständnis für diejenigen aufbringen, die dies alles als Horror-Szenario von sich weisen und die guten alten Verfahrensweisen loben. Derartiges Verhalten führt aber nicht weiter, und die Wege zurück hatten selten eine Zukunft. Es ist also müßig zu fordern, daß Bibliotheken mehr Personal brauchen, um die digitalen Dokumente in unseren bewährten Katalogen zu verzeichnen. Daß dies eine dumme Utopie ist, belegt folgende überschlägige Rechnung. Bei bibliothekarischer Erschließung wäre gegenwärtig – bei einer jährlichen Steigerungsrate von 100% – ein Aufwand von mehr als drei Milliarden US-Dollar nötig, um alle 80 Millionen HTML-Seiten zu katalogisieren; und das bei einer durchschnittlichen Verweildauer von Web-Seiten im Internet von rund 80 Tagen.[4] (Bei Beschränkung auf eine Million Web-Seiten fielen immerhin noch Kosten von 45 Millionen US-Dollar an.) Selbst der schnellste Geschäftsgang in großen Bibliotheken würde – die unvorstellbar große Summe Geldes als verfügbar vorausgesetzt – stets an historischem Material arbeiten. Eine groteske Idee; man muß dies aber wirklich so konkret zu Ende denken, um manche Ansätze in ihrer Unsinnigkeit zu entlarven.

Schließlich sind auch die Möglichkeiten einer intensiven Analyse und Aufbereitung von Rechercheanfragen, die sich zu einer Vielzahl von Suchen sogar in mehreren Sprachen weiterentwickeln lassen, mit ins Kalkül automatischer Erschließung zu ziehen.

Diesen Gedanken abschließend erscheint es wenig hilfreich, darauf hinzuweisen, daß die intellektuelle Erschließung der automatischen qualitativ überlegen ist. Derartige Fragen sind falsch gestellt. Richtig formuliert, heißt die Frage: Ist es uns lieber, für bestimmte Informationen gar keinen Nachweis zu liefern als die Ergebnisse automatischer Verfahren als Sucheinstieg anzubie-

---

4 *Kahle/Lyman*: Archiving Digital Cultural Artifacts (wie Anm. 1).

ten? Die Antwort kann nur dahin gehen, daß wir automatische Verfahren klug in unsere Erschließungskonzepte einbeziehen und sie möglichst sachnah an die jeweilige Klientel herantragen müssen.

Management der Heterogenität, Interoperabilität

Unter dem Eindruck der Möglichkeiten der Erschließung digitaler Medien und nach Reflexion der Rückwirkungen auf die traditionellen Erschließungskonzepte steht wohl zweifelsfrei fest, daß neben die immer stärker reduzierte Eigenleistung mehr und mehr die Übernahme von Fremdleistungen sowie die Einbindung automatischer Verfahren treten werden. Sinnvoll ist dies jedoch nur, wenn die Übergänge erleichtert werden, wenn wir die Angebote untereinander vernetzen. Management der Heterogenität einschließlich der Interoperabilität der Systeme und Systemkomponenten ist die Aufgabe der Zukunft, nicht die Harmonisierung in geschlossenen Systemen.

Schluß

Fassen wir zusammen: Die Verunsicherung in der gesamten Szene der wissenschaftlichen Information und Kommunikation ist aufgrund essentieller Änderungen auf vielen Gebieten mit jeweils erheblichem Ausmaß groß. Dies betrifft zwar alle Medienarten; insbesondere sind es aber die digitalen Medien und die elektronischen Vertriebswege, die an vielen Stellen eine Neukonzeption nahelegen. Dies ließ sich in den zwei näher dargestellten Bereichen belegen, die in der Vergangenheit zwar im Detail ständig Anpassungen erfuhren, im Grundansatz der Aufgabendefinition aber seit vielen Jahrzehnten Bestand haben. An die Stelle der Erwerbung im klassischen Sinn tritt das umsichtige Ressourcenmanagement, die Katalogisierung findet sich im vielfältigen, flexiblen Erschließungsmanagement wieder.

Die aussichtsreichen Perspektiven liegen auf Gebieten, die in der Vergangenheit vernachlässigt wurden, weil es keine praktikablen Instrumente gab. Nun, da wir über die Technik verfügen, benötigen wir reale Lösungen. Es lohnt, die Herausforderungen aktiv anzunehmen und vorrangig nicht die Gefahren, sondern die Chancen zu sehen. So betrachtet, waren die Zeiten für Bibliotheken noch nie so gut wie heute.

Strategien und Kooperation bei der Digitalisierung von Archiv- und Bibliotheksgut

Die Verteilte Digitale Forschungsbibliothek als Infrastrukturförderung der Deutschen Forschungsgemeinschaft

Von JÜRGEN BUNZEL

Introduction

The Deutsche Forschungsgemeinschaft has in the area of Digital Library Development various project activities and funding concepts.

I will start with a short explanation of the

- structure of our funding programs and the description of the approach and vision of the Distributed Digital Research Library.

I will then give a short overview of ongoing project activities sponsored by the DFG in the areas of

- Technology Development for Digital Libraries,
- Retrospective Digitization of Contents for Digital Libraries, and
- Integrating electronic publications into Digital Library Services.

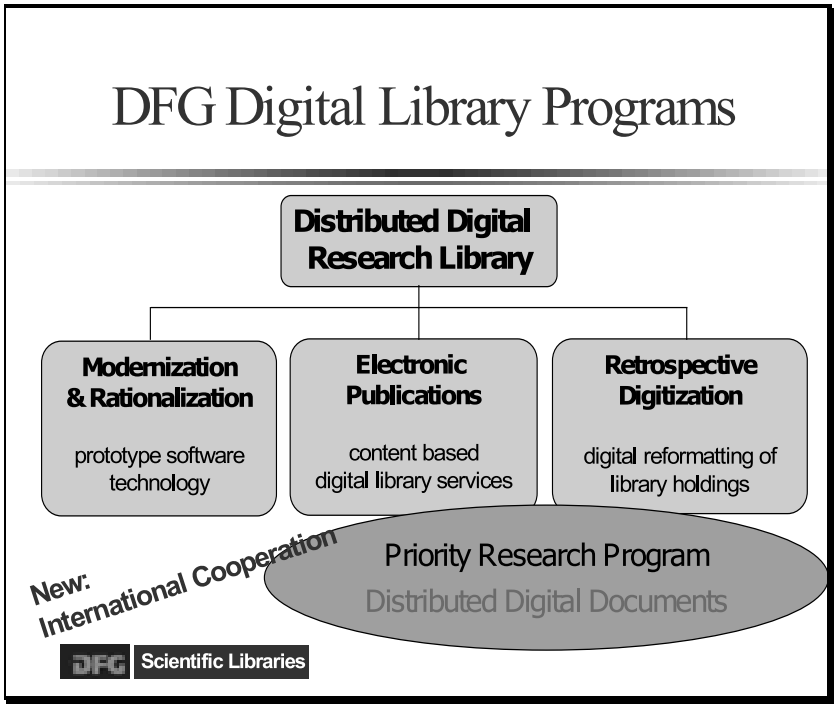
Finally, I want to make some remarks on

- Perspectives for international cooperation.

The DFG Digital Library Programs

In the last two years the Deutsche Forschungsgemeinschaft has redesigned and extended its program structure with respect to the new demands brought about by electronic information networks and multimedia publications.

One basic step has been the establishment of a new priority area of project funding for scientific libraries, an action line, which we call the *Distributed Digital Research Library*. This action line has officially been started at the beginning of 1997.



It comprises three distinct funding programs, which are closely coordinated and integrated under the new heading.

The first program is called *Modernization and Rationalization in Scientific Libraries*. It is aimed at the development of pilot technology, especially software developments of precompetitive nature. At the moment there are about 20 projects in progress in this program.

The second program is a funding program for *Electronic Publishing*, which is targeted on pilot development of services for the provision of electronic publications and multimedia in scientific libraries. So while the first program, I mentioned, focuses on technology, this one is primarily focused on contents. Here also we support about 20 projects at the moment.

The third component of our funding concept for the *Distributed Digital Research Library* is the Retrospective Digitization. Last year we were able to set up a separate, new funding program for this task, for which we received a considerable additional budget. Today, in the second year of this new program, there are 34 digitization projects in progress.

Basic Research on Digital Libraries

An important milestone in reshaping our library programs has been a newly established close cooperation with our computer science department.

In 1996 a group of computer scientists proposed a five year priority program for basic research on *Distributed Management and Handling of Digital Documents* to the DFG.

During the preparation of this priority program the libraries group of the DFG and our computer science department have joined forces to bring computer scientists in close contact with librarians, so that librarians would transmit their requirements to reseachers, and researchers would be provided with real-life test-beds for their solution concepts.

This priority research program is now running in it's second year with about 25 projects funded. In eight of these research projects libraries are involved as managing partners.

International Cooperation

With the beginning of next year we will extend our activities for Digital Libraries with another component, namely: financial support for international cooperation.

We will be able to provide about half a million DM additionally and exclusevily for this purpose. Applications can be made for travelling expenses, but also for additional staff or equipment needed to establish project cooperations with partners abroad. Prerequisites are

- that the basic project, which shall be extended by international cooperation, is funded in one of the program areas of the distributed digital research library, and
- that there is a foreign partner who actively takes part in the project. Ressources for the foreign partners shall not be provided by DFG but have to be provided by foreign funding agencies.

The Distributed Digital Research Library

In total there are presently about 90 projects on digital library development running in this coordinated program context of the Deutsche Forschungsgemeinschaft. Our budget for these activities has been about 8 Mio DM in 1997. This year we will provide already more than 12 Mio. DM, and I am expecting increasing budgets for the years to come.

Let me point out two essential aspects concerning the basic philosophy of our concept.

1. We do not regard Digital Library Development as a self contained scientific research, but as a contribution for building up modern information infrastructures to support science and research. Even the Priority Research Program on distributed documents, beyond being basic research in Computer Science, is expected to lead to practical applications improving information infrastructure for science.
2. The distributed digital research library – in the first place – is a funding concept, i.e. a set of interrelated and coordinated funding programs of the Deutsche Forschungsgemeinschaft. It is not a top-down concept for setting up a distributed digital research library in an organizational sense.

I want to stress that every library in Germany or every scientific working group can freely come up with its own ideas and propose project applications in the three funding programs. There are basically no restrictions concerning project contents or project methodologies. All project proposals are evaluated only in terms of quality.

Our basic philosophy is to foster a broad variance and competition between different approaches to establish technologies, contents and methods for digital libraries.

Of course, in a longer perspective we would appreciate if results were transferred and taken up between different projects. And we would like to see digitial collections growing together into more comprehensive and interrelated service offerings.

Thus, a *Distributed Digital Research Library* in an organizational meaning may and shall emerge from these project activities. But this can only grow buttom up from a multitude of diverse approaches, and not by a top down planning approach.

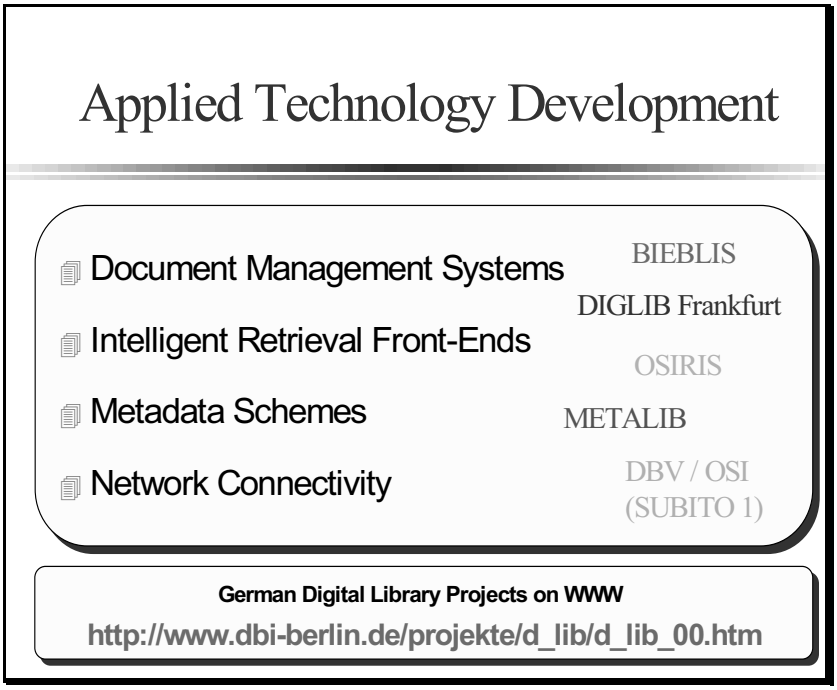
Applied Technolgy Development

I will now turn towards the first building block of our strategy: The innovative applied technology development for Digital Libraries.

One major project activity, here, was to provide libraries with document management systems to integrate digitized collections and electronic publications into their Online Public Access Catalogues.

Digital Document Management Systems

One of the first fields in this area was the project BIEBLIS done by the University Library of Bielefeld. It has been successfully completed in 1997. BIEBLIS is a document management system, which provides very flexible schemes for the seamless integration of digitized fulltext-sources into an internet based library OPAC. Meanwhile the results of the project have been licensed to a Commercial Provider. So the system can be bought. It will provide the technical basis for digital library project of North-Rhine-Westfalia and it will be used in some of our projects on retrospective digitization.



A second, quite advanced project is done at the Institute of Computer Science in Frankfurt. They are developing a Digital Library solution, which heavily relies on software agents as the basic infrastructure element for integration and federation of document types. A prototype of this system has been used for the internet presentation of digitized political pamphlets of the 1848 revolution, a digitization project funded at the University Library of Frankfurt.

In addition the two Service Centers for Digitization in Göttingen and Munich will set up prototype document Management-Systems in cooperation with commercial software providers. So there should be quite a few technical solutions for libraries which they can choose from, to integrate direct user access to digitized books and journals into their electronic service environments.

A second field of project activities in technology development are

Intelligent Retrieval Technologies

It is common understanding that precision and recall of information retrieval have to be significantly improved if the user shall not get lost in the desert of digital information space. In the last few years we had a couple of project proposals addressing these problems.

One project already completed with quite promising results is OSIRIS.

OSIRIS is a project on intelligent retrieval technologies, that applies really sophisticated computer linguistic techniques. It was done at the University Library of Osnabrück.

Other projects on real life applications of intelligent retrieval techniques are running at the University Library of Düsseldorf and Oldenburg.

Most of these projects originated from problems with information retrieval in electronic library catalogues. I hope that their results will also contribute to improving access modes for digital objects in general.

Metadata Schemes

A third major theme in applied technology development for Digital Libraries are Metadata Schemes.

Major German libraries have set up a joint core project on metadata issues. This multi-partner project called *MetaLib* has been started in August 1997. It is coordinated by the Deutsche Bibliotheksinstitut and consists of three parts:

The University Library of Göttingen is responsible for collecting the technical know-how on metadata standards to be able to take part in the international standardization efforts and to implement advanced application systems.

The Deutsche Bibliothek will be responsible for transfering results from the new metadata developments to the field of traditional cataloging standards and bibliographic formats.

The Bayerische Staatsbibliothek in Munich will study different metadata requirements for different subject communities or different material types and develop appropriate tools for transparent access to heterogeneous metadata sets.

I think that this project has already made significant contributions in strengthening the voice of German librarians in the international metadata discussions. In addition, it will be essential that metadata are practically applied and used in the digitization projects.

Z39.50 Network Connectivity

One of the most important technical prerequisites for digital libraries is the network connectivity of information systems. We have been very active in this field in Germany since the late 1980s. One of the most prominent activities is the large scale project on *German Network-Interconnection of Library- and Online-Information Systems* or with its somewhat out dated acronym: DBV/OSI.

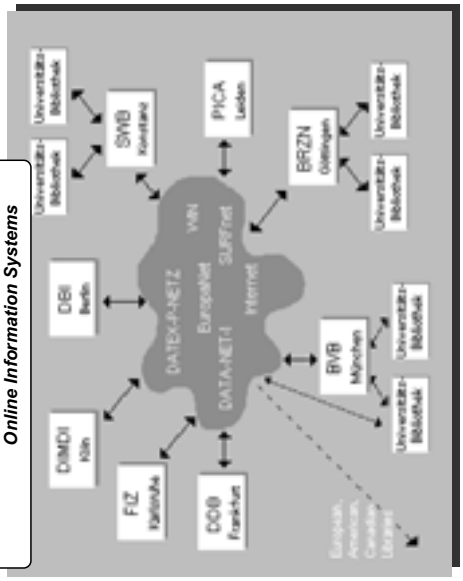
The main goal of this project was to establish network interconnections between selected regional library systems in Germany and two major Online-Information providers. The network technology has been based on the Z39.50 standard. This project was jointly financed by the DFG and the Federal Ministery of Research between 1993 and 1997. It has been successfully completed at the end of last year.

So on the technical side we have now functioning and heavily used Z39.50 interconnections between the different library systems in Germany. This network structure is the technical basis for SUBITO 1, the new nationwide electronic document delivery service of German Libraries, which has been opened last november.

I hope that this networking structure of Libraries will be extended for electronic document delivery and direct access to digitital collections. This would provide a sound technical basis for the *distribution* component of our vision of a digital research library.

DBV / OSI Project

DBV / OSI = Z 39.50 Network of Library- and Online Information Systems



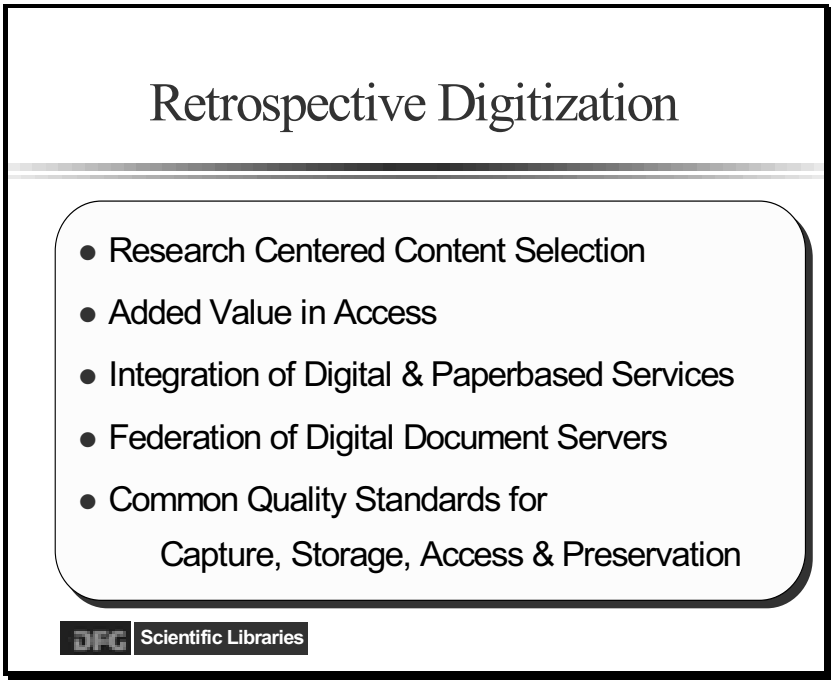
DFG Scientific Libraries

DBV/ OSI connects

- **Regional Union Catalogue Systems**
- **Regional & Local Systems**
- **Libraries & Online-Providers**

Technical basis for SUBITO

- **Z 39.50 Item Order**



Retrospective Digitization

Let me now turn to the issue of providing content for the distributed digital research library. As I have already mentioned, there are two DFG funding programs targeted on this area

- Retrospective Digitization of paperbased library holdings
- and the Integration of electronic publications, that are documents *born digital,* in to library services.

It seems to me that in Germany we have taken a somewhat different approach to retrospective digitization compared to the United States, Australia or France. Most of the countries that I know, have started with digitizing central monuments of their cultural heritage.

Our approach is lead by the following principles:

Service value for researchers
is our primary goal. We want to provide library collections in digital form, which are on the one hand urgently needed by researchers and on the other

hand very rare, valuable or not easy to obtain in their traditional paperbased form.

Added-Value in access
is an issue very much stressed by the evaluation committee for digitization projects. The opinion is that digitization should provide a marked added-value in access to document contents beyond the paperbased document formats. This e.g. can be provided by full-text search, improved automatic or intellectual indexing of contents and interlinking of content indexes for different collections.

Seamless service integration of digital and paperbased libraries
All materials that are digitized should be catalogued and indexed in the same databases as the paperbased library holdings, so that all types of holdings shall be transparently accessible in the local and regional union catalogues.

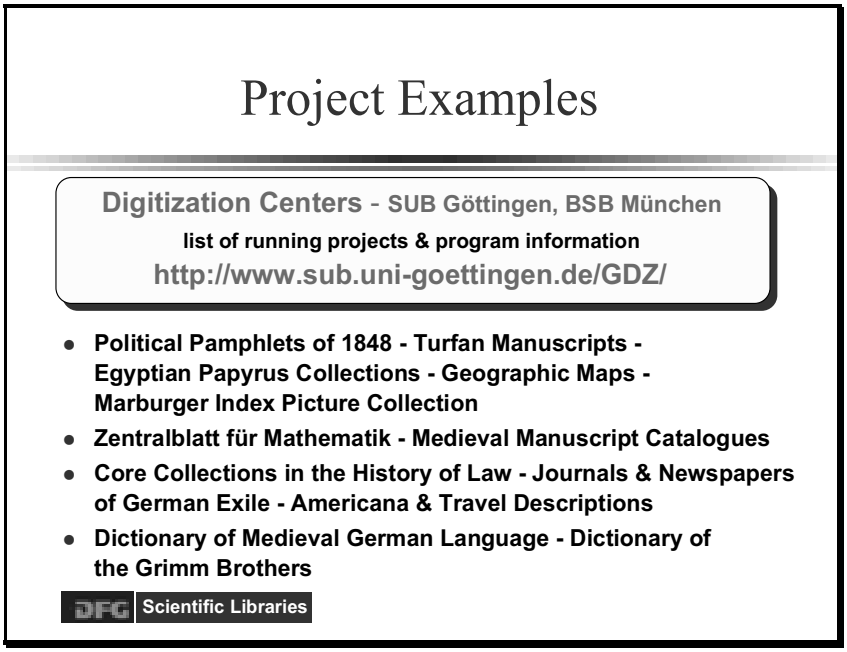
Federation of distributed digital collections
This means integrating the diverse digitization activities into one unified and comprehensive information system based on federated, not necessarily homogeneous technical systems. In the longer run, the distributed digital research library should really become visible on the net as one virtual reading room, where real library users will be able to go to the real digital documents on the virtual shelves, that 7 days, 24 hours a day.

Common Quality Standards for Capture, Storage, Access and Permanent Archiving of Digital Objects
In preparatory projects done by the University Library of Göttingen, technical standards and guidelines for retrospective digitization have been established, which we recommend for DFG funded projects.

Project Examples

The DFG funding program for retrospective digitization, which started last year, has received a very good and active response by libraries and scientific institutions. We had a lot of high quality project applications covering a broad range of subjects, material types, and technical approaches. I regard this as a benefit of such a highly experimental area where almost no established experiences exist.

To assist libraries in preparing and running their digitization activities we have set up two centers for digitization at the University Library of Göttingen and the Bayerische Staatsbibliothek München. They are responsible for gathering technical know-how on digitization issues and disseminate it to libraries.



To give you an impression of the diversity of projects themes covered in the program I will give a short overview proceeding by categories of material type.

Quite a few projects digitize single page materials, e.g. political pamphlets of the 1848 revolution are digitized at the University Library of Frankfurt. Other projects of this type are the digitization of single page oriental manuscripts and egyptian papyrus collections.

Then we have projects concerned with geographical maps, which pose very special technical problems and projects concerned with picture collections. The most prominent here is the large scale project on digitzing the German National Picture Archive of Arts and Architecture, the *Marburger Index*, with more than one million pictures.

Then we have projects concerned with geographical maps, which pose very special technical problems and projects concerned with picture collections. The most prominent here is the large scale project on digitzing the German National Picture Archive of Arts and Architecture, the *Marburger Index*, with more than one million pictures.

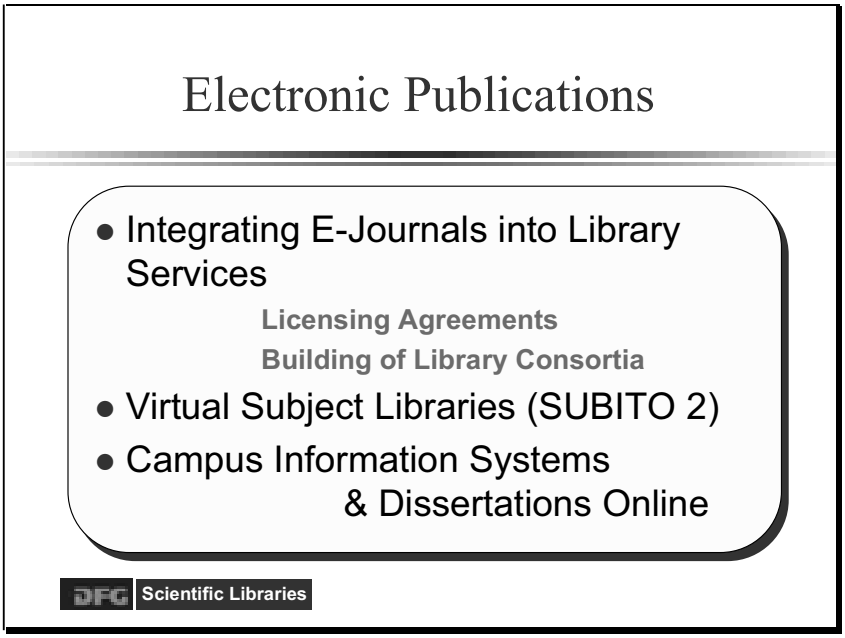
A second category of projects is the digitization of bibliographical reference sources, which entries are linked to digitial facsimilies of the journal articles or books they refer to. One of these projects concerns the Zentralblatt für Mathematik, which is the core bibliographical resource of 19th century ma-

thematics. We have taken steps for another project of this type, in which our existing catalogues of medieval manuscripts will be linked to digital images of the manuscripts themselves.

The third category are projects concerned with the complete digitization of medium sized collections of books and journals. For example core collections of the history of law at the Max-Planck Instuitute at Frankfurt, the digitization of journals and newspapers of the German Exile between 1933 and 1945 by the Deutsche Bibliothek and a comprehensive collection of 19th century American literature and travel descripions at the University Library of Göttingen.

A final type of activities touches the borderline between retrospective digitization of library holdings and scientific editions. We have quite a few digitization projects for central dictionaries of german literature, e.g. the Dictionaries of medieval German language and the famous Dictionary of the Grimm-Brothers. Also the full-text digital conversion of selected texts of medieval latin writers, done at the University Library of Mannheim, belongs to this category.

Digital Libraries and the DFG-System of Special Collection Fields



Finally I want to make a few remarks on our funding program for integrating electronic publications, i.e. publications born digitally, into the library environment.

For the Deutsche Forschungsgemeinschaft a key issue in this context is the effect of Electronic Publishing on our system of special collection fields and its future development in the digital age.

Most of you will know the basic idea of this system, which in German is called *Sondersammelgebiete*. As a substitute of a classical National Library the complete subject range of science and the humanities is divided between more than 20 libraries, each of which takes care for the specialized literature collection in its assigned subject fields.

The *Sondersammelgebiete* have always been the core of DFG library funding. We spend nearly 50% of our budget on the acquisition of books, journals, and CD-ROMs in the *Sondersammelgebiete*.

One of our main concerns is how to transpose this system and its underlying ideas into the digital age.

Project activities on extending digital services into the special collection field system have started this year. There are two main action lines.

Integrating Electronic Journals into the System

An issue of special importance for the DFG is a strategy to guarantee an open access to electronic journals for users of the special collection system.

Traditionally the special collection system is used by interlibrary loan, which is open for an unrestricted, not institutionally bound user community.

It will be difficult to contract free interlibrary loan permissions for electronic delivery of digital content. So one possible solution for sustaining unrestricted information access in our special collection system could be pay-per-view schemes or pre-purchased fixed contingents of journal article accesses. We have just started a couple of experimental projects on the procurement of electronic journals in our special collection system to test appropriate licensing agreements and the effects of library consortia.

Virtual Subject Libraries

A second main action line in this program are Virtual Subject Libraries. Virtual Subject Libraries are a concept developed within the SUBITO II initiative. The DFG has adopted this concept as a primary concept for the special collection field going digital.

The idea is that the special collection field libraries become the essential crossing points in the network, which

- initiate and coordinate the collection of web-ressources, which relate to their respective subject field,
- organize cooperative projects for cataloguing and indexing of these ressources,
- provide specialized search machines,
- and perform aggregator functions for access to electronic journals and other types of electronic publications.

You can already see this concept in action, in two forerunner projects for geosciences and mathematics. Both projects are run by the University Library at Göttingen.

There will be a couple of additional pilot projects on virtual subject libraries. The following subjects are discussed for selection: social sciences, engineering sciences, Anglo-American culture, history of early modern times and performing arts.

Campus Information Systems & University publications

Other project activities in the program for electronic publications are pilot developments for campus information systems and the problems related to electronic university publications.

A core project here, which has started this year, is Dissertations online. This project is done in cooperation of several universities with a group of scientific societies. It shall provide solution concepts for all the technical, institutional, organizational and legal regulations necessary to set up a functioning system for the management of electronic dissertations.

International Cooperation

One of the intentions of this conference is to promote new partnerships between German projects on Digitization and Digital Libraries and related activities abroad.

I am very pleased that we will have the opportunity to talk about that with our American Guests this afternoon in a meeting of our evaluation committee for the retrospective digitization program.

Of course, international cooperation is not restricted to our American partners. The international funding component for the Digital Library programs of the DFG, which I mentioned at the beginning, is open for partnerships with all countries. And we are are eager to intensify working contacts with our European guests from Belgium, Switzerland, Austria and the Netherlands, especially with regard to the 5th framework program of the European Union,

which will contain an action line for preservation and digitization of cultural heritages of European museums, archives and libraries.

When thinking about cooperative projects, we should scan the whole spectrum of joint interests.

This might concern cooperation of methodological issues, which have played a major part in this conference, as well as the joint development and exchange of technical solutions and software systems.

Perhaps, it would be interesting to think about cooperation between Digitization Centers, which e.g. could pool their knowledge bases and provide advice and assistance through an international network of newsletters or helpdesks.

Naturally, a key focus of international cooperation should be content based services, i.e.

- joint digitization projects for related collections, and
- the interlinking of web-sites, content repositories and access databases to take first steps towards internationally distributed digital libraries.

In my opinion, one of the most interesting and most important issues, is, that libraries and archives play an active role in setting up international collabora-

tories of researchers and scholars, in which the provision of digitized materials builds the basis for internet-based research cooperations using all facilities from electronic mail, preprint servers, virtual experiments and networked data analysis up to virtual conferences and virtual classrooms.

I think these kinds of collaboratories really embrace the concept of a digital *research* library, according to our definition.

One of the main factors of success factors of all our digital undertakings will be that our service offers gain user recognition. Not information, but user recognition is the dramatically scarce factor in the internet world. Perhaps user recognition will be even more important than user acceptance.

Certainly one of the most promising strategies of gaining user recognition is to actively engage in the organization of ones own user communities. This is exactly what can be done with digital collaboratories. And it is very similar to the strategy traditional research libraries, like for example the Herzog August Bibliothek Wolfenbüttel, succesfully employ for many years: actively organize ones own user communities around ones collections.

Each digitization project has to be well aware, which are the user communities it wants to serve. The most promising projects will be those which play an active part in shaping an international user community for a digital collection.

Council on Library and Information Resources Vision and Future

Von Deanna B. Marcum

Many of you in this room know Hans Rütimann very well, and you have benefited from the wealth of information he circulates between the United States and Germany – indeed, around the world. You have known him as the International Program Officer of the Commission on Preservation and Access, and wonder why he and I now talk about CLIR – *The Council on Library and Information Resources*. My role today is to explain two things: (1) What CLIR is – its programs and directions, and (2) how such an organization moves into specialized efforts such as the Digital Library Federation when the need arises.

History

As a historian, I am compelled to go back in time to explain cause and effect, so allow me to provide a quick review as context. The Council on Library and Information Resources was officially created in 1997, through the merger of two predecessor organizations – the Council on Library Resources (CLR) and the Commission on Preservation and Access (CPA). The CLR´s roots go back to 1956. The CPA was formed in 1986. To understand what today´s organization is, one must first look at these parent organizations. Re-creating the CLR would probably be impossible today. Neither government nor professional organizations provided the impetus for its formation. CLR was born of the passion of one individual, Louis B. Wright, who was a scholar/librarian.

The 1950s were a period of explosive library growth. Wright believed that university libraries were competing with one another and, in the process, expending large sums of money unproductively. He was particularly concerned about the amount of money being spent by individual institutions to microfilm European documents, with no plan to make the copied documents available to a wider audience. He began to wonder why some organizations could not coordinate the microfilming efforts to avoid costly redundancy.

The advent of new library technologies, the rapid expansion of libraries, and the proliferation of individual, uncoordinated activity deeply concerned Wright. He believed these problems could be addressed only by an overarching organization built on intellectual integrity and commitment to scholarship. Wright turned to his longtime friend, Fred Cole, who was the program

officer for education at the Ford Foundation. He convinced Cole to fund a conference of fifteen scholars (including humanists and scientists), publishers, and librarians to consider what kind of coordinating body was needed. The conference, held in January 1956, led to a more focused discussion of library matters, with a much expanded audience, in March of the same year. Out of that March meeting came a proposal to the Ford Foundation calling for the establishment of an independent, non-membership organization of sufficient stature to address the problems faced by the library community in a period of escalating growth and competition. In May 1956, the Foundation responded with a grant, and the first meeting of the CLR Board was held in September 1956. Can you imagine creating an organization in such a short time today?

The first Board of the Council on Library Resources was constituted, for the most part, of Wright's close friends who were also scholars and documentary editors. The single demand made by the Ford Foundation was that the chairman of the Board must be a businessperson. Gilbert Chapman of the Yale Lock Company was named chairman. The first president, the individual who managed the organization, was Verner Clapp. Clapp had been for many years the Deputy Librarian of Congress. His interests were broad and international. He was not a trained librarian, but a longtime LC employee vitally interested in bibliographic control, the use of technology to improve library operations and international cooperation. Deeply disappointed that Quincy Mumford, and not he, was named Librarian of Congress after Archibald Macleish resigned, Clapp used the post of CLR president to broaden to other institutions the reach of programs he had started at the Library of Congress. He took seriously the opportunity to develop a center of intellectual activity that would examine the role and function of the library.

The early programs of the Council on Library Resources focused on bibliographic structure, automation of library operations and preservation. There was also a separate set of international activities, mostly aimed at helping European libraries recover after the devastation suffered during World War II.

Preservation has always received primary attention. Among the very first grants made by the Council in 1957 was a large award to the Barrows Laboratory, part of the State Library of Virginia, to study the causes of paper deterioration. This concern about preservation of books and journals continued through the first two decades of the Council's history, with many of the grants going to individual libraries to help them develop local preservation programs.

When Warren J. Haas was named president of CLR in 1978, he was the Librarian at Columbia University. Faced with a massive, deteriorating collection of nineteenth century scholarly materials, he called for a national plan to preserve the materials that were literally turning to dust on the shelves of the large research libraries. He knew that individual efforts, while important, would not result in a preserved intellectual heritage unless someone provided leadership for a national effort. He asked the Association of American Universi-

ties to join the Council in tackling the problem. There was first a joint task force and later a special committee made up of librarians and scholars to study the brittle books problem and to make recommendations for collective action. In 1985, the special committee urged the formation of a specialized, highly focused organization to address the issues. In 1986, the Commission on Preservation and Access was created as a separate organization, with the sole mission of improving the prospects of preserving at least a significant portion of the books that were disintegrating in the nation's libraries.

Under the inspired leadership of Patricia Battin, the Commission on Preservation and Access, working closely with the National Endowment for the Humanities (NEH), developed a national strategy to undertake massive reformatting – that is to say, microfilming – projects in the major research libraries. The Commission established a technical advisory committee to make recommendations on appropriate technologies and an advisory committee of preservation administrators. Targets for microfilming production were established and reported on annually to the Senate oversight committee responsible for the NEH budget. The progress in the early years, especially, was astounding. There were the predictable debates about which collections should be filmed first, about filming selected items rather than entire collections, and, as you might expect, about which institutions most deserved federal funds for microfilming projects.

Today, I am often asked if the new organization, the Council on Library and Information Resources, continues to be deeply interested in preservation and access. Or, have we moved entirely into the digital library realm? The answer is that, even during the late 1980s through the mid-1990s, the Commission on Preservation and Access, while focusing its laser-like attention on the problem of deteriorating books and journals, was beginning to recognize the changes that would result from the emergence of digital technology. It was a group of librarians working on projects to test the feasibility of using digital technology for preservation that urged first the Commission on Preservation and Access and later CLIR to coordinate what became the Digital Library Federation. The Federation grew out of informal discussions among a group of eight librarians (called the LaGuardia eight in honor of the meeting site – LaGuardia Airport). The group grew to 12 institutions which concluded that they needed to look at the broader implications of the technology. The group drafted a document entitled the *Digital Preservation Consortium* mission statement. Members called for a planning strategy for the development of digital libraries, and began to organize themselves to continue their local efforts while also informing one another about their findings.

At about the same time that the group issued its document, the Library of Congress issued a press release announcing its intention to create a national digital library. Since the 12 institutions of the consortium, all major research institutions, were themselves establishing digital library projects, they wanted assurance that what they were doing and what LC was doing would be com-

patible. The group approached the Library of Congress as well as the National Archives to join a new effort – the National Digital Library Federation, soon to be more simply called the Digital Library Federation, or DLF.

This sequence of events took place just prior to my dual appointment as president of the Council on Library Resources and the Commission on Preservation and Access. The boards of these organizations realized there were several reasons to combine the two: the organizations? mission and goals could be sharpened, the staffing streamlined, and money saved. The Digital Library Federation was an excellent example of how distinctions between the two previous organizations had become blurred. The DLF was needed by only a few institutions that were deeply involved in digital projects, and was, therefore, not a good candidate for a professional association initiative. The coordinating role was lodged with CLIR so that DLF institutions could learn from one another efficiently.

Each of the current 22 members of the Federation pays a significant fee – $ 24 000 per year – to CLIR, with which we hire a DLF staff – a director and research associate – who gather information from all participants, develop projects that respond to the most important unresolved questions, and report results to the broader library and archival communities. We expect this organization will be needed only a few years. Within five years, the work now being done by DLF will be commonplace; there will be no special role for CLIR and we shall be working on something new. That is the nature of this catalytic organization.

What then, does CLIR mean by preservation and access? The answer is quite simple: preservation and access are at the heart of everything we do. It is our conviction that libraries and archives have the societal responsibility to preserve information resources in whatever form they exist and pass them securely and reliably on to the next generation of scholars and students. Digital technology changes some of our considerations about processes and techniques, but the concern remains central.

CLIR recognizes that an effective preservation strategy relies on a variety of approaches. The Brittle Books program that emphasizes microfilming major collections in research collections continues with NEH funding. Although the NEH budget has been reduced over the last few years, there is an annual allocation of approximately $20 million that is distributed mostly to research universities and archives to allow them to microfilm, in the national interest, collections deemed in jeopardy of being lost forever. CLIR continues to stress the importance of environmental conditions in the care of collections and in preventing paper deterioration. It has funded work at the Image Permanence Laboratory at the University of Rochester to develop a model that institutions can use to establish ideal environmental conditions for long-term preservation of different formats of materials. Additional funding is needed to bring this model to full operation, but we continue to pursue the goal.

It is the concern about the persistence of information over time that leads us into the digital arena. Much scholarly communication in the scientific and technical fields exists only in digital form. Our discussions with scholarly societies have led us to believe that even in the most conservative areas of the humanities, journals will increasingly become electronic. If libraries are to continue their important traditional function of acquiring and preserving the significant intellectual record, then we have no choice but to find ways to make those electronic documents persist into the future.

This realization has led to several Commission and, later, CLIR, projects. We remain aware, always, of the importance of sound managerial decisions. Acknowledging that microfilm is a reliable medium for capture and storage, we also recognized that it is a poor presentation medium. We began to inquire about the possibility of digitizing preservation microfilm. Is it feasible? Can it be done at acceptable costs? To answer these questions, we helped support Cornell and Yale Universities to learn more about converting digital resources to microfilm and microfilm to digital. To better understand the results of these efforts, collectively called Project Open Book, CLIR formed a study group in spring 1998 to plan a hybrid conversion approach to preservation and access. We are working closely with the National Endowment for the Humanities and the Research Libraries Group in this endeavor, which aims to codify best practices in hybrid conversion from microform to digital images. The project's first product will be a paper that summarizes the current state of knowledge, identifies areas that need further development and research, and recommends a course of action to meet those needs. CLIR will distribute the paper in January 1999 and invite comment from the preservation, vendor and technology communities. This will be followed by a period of collaborative work on the areas of immediate need. The final paper will be issued next summer and will include published guidelines and best practices that the federal agencies may require in preservation reformatting projects receiving government support.

Within the Digital Library Federation, a number of projects are under way that will shape our approach to digital preservation. Earlier this year we commissioned Jeff Rothenberg, a senior research scientist at the Rand Corporation, to write a paper on a recommended strategy for digital preservation. We have received his report entitled *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Rothenberg argues that the long-term digital preservation problem calls for a long-lived solution that does not require continual heroic effort or repeated invention of new approaches every time formats, software or hardware, document types, or record-keeping practices change. He finds serious deficiencies in the migration techniques that are most often used by libraries and archives today. He argues for the emulation of obsolete systems that can be used on future, unknown systems to run a digital document's software. Though it may not be feasible to preserve every conceivable attribute of a digital document in this way, it

should be possible to recreate the document's behavior as accurately as desired – and to test this accuracy in advance. CLIR will soon publish the report, though it is certain to be controversial. Our next step will be to convene a group of prestigious computer scientists to assess the feasibility of the approach Rothenberg suggests.

In the meantime, because so many libraries have already embarked on a migration strategy, we continue to support projects that help us understand better the requirements and limitations of that strategy. We have funded Cornell University Libraries to develop a risk assessment strategy for the migration of digital data from one system to another. The final report will be published late next spring.

The Digital Library Federation has identified a doctoral student at Carnegie Mellon University who has developed a network-based service for converting digital files from one format to another. Several libraries within the Federation are joining forces to determine if the prototype can be turned into an operational service that would allow all libraries to use a centralized service for converting documents on an as-needed basis.

This list is not exhaustive by any means, but it gives you an idea of the kinds of projects that CLIR has included in its agenda. They are hard to categorize – they have a preservation motivation, even when they appear to be better described as digital library projects. And that is precisely our dilemma. It is no longer possible to categorize the work that desperately needs to be accomplished. CLIR is a model of the kind of organization that will become more commonplace, I predict. We have an independent Board, but we lack both a membership base and an endowment. We raise from private sources – mostly from large U.S. foundations – the money needed to carry out our agenda. We are in a position of always having to argue our case. We must be in close enough touch with the library community to be so bold as to speak for it. The only authority we have is that ceded to us by our colleagues in the field. The four areas we have chosen to emphasize – preservation and access, digital libraries, economics of information, and leadership – all have international dimensions, too. Increasingly, we recognize the need to engage partners internationally to carry out the agenda. Much has already been done internationally in the preservation and access program, thanks to significant funding from the Andrew W. Mellon Foundation for that purpose. Digital libraries offer an exciting and timely opportunity to combine our efforts with the substantial efforts that have been made by other countries, as well.

The new CLIR is an organization that takes the best of the two predecessor organizations. CLIR continues to be opportunistic; that is, it tries to be always alert to possibilities for action on a particular issue. CLIR is a catalytic, convening organization. It tries to find the best minds to consider issues of broad concern. It remains an independent, neutral body – one that can bring together disparate organizations and individuals to try to resolve deep-seated problems. And it tries to keep its focus on the truly important issues of natio-

nal and international concern. CLIR's vision is that university libraries must transform themselves if they are to continue to serve the needs of scholars, researchers and students. They must work as effective partners with other managers of information resources on the campus and external to the campus. They must adopt leadership styles that depend much more on collaboration than on hierarchy. They must learn much more about digital technology in order to apply the most meaningful and cost-effective solutions to the problems of higher education. And, as some problems are resolved and solutions are rendered operational, CLIR will be assessing what needs to be done next.

Better Access to Primary Sources 1983–1998

Von JOHN W. HAEGER

In the dark days of the Second War, the French historian Marc Bloch – best known for a landmark work on *Feudal Society* – driven from academe and serving as a leader in the Resistance at the age of 55, began a manuscript that he never finished. Alas, he was executed by the Nazis in an open field near Lyon ten days after the allied landing on the beaches of Normandy in 1944. Bloch called his manuscript *Apologie pour l'Histoire ou Métier d'Historien* – later rendered in English as *The Historian's Craft*. It's the sort of book academics often describe as a *think piece*, the kind of thing you can write without sources, a little book physically. If *The Historian's Craft* is less significant than Bloch's synthetic work or his seminal monograph on medieval peasants, it is still remarkable in many ways. Bloch talks about the way historians work and the kind of history he thought good historians should write. He quotes the poet Paul Valery, who cites the impact of electricity as example of a hugely important and influential phenomenon which must of necessity elude the historian, because, according to Valery, there are no documents which refer to it specifically. *Who believes,* says Bloch in response, *that the electrical companies have no archives, no records of consumption, no charts of the enlargement of their networks?* The truth, he says, *is that until now historians have simply neglected to question these documents. There is no reason to make history as it can be the scapegoat for the sins which belong to bad history alone.* Discussing evidence, Bloch talks about *unwritten evidence* and *vestiges of the past.* Knowledge of human activities past, he says, is a knowledge of their *tracks.* Sometimes inadvertent tracks made by w*itnesses in spite of themselves.* The variety of historical evidence, he reminds us, *is nearly infinite. Everything that man says or writes, everything that he makes, everything he touches can and ought to teach us about him.* Who, Bloch continues, could pretend to describe the history of land use, if he were not equally well-equipped to read critically a medieval charter, to explain the etymology of place names, to date the ruins of dwellings, and to analyze the plant life proper to pasture, field and moor? And these are, of course, the hallmarks of Bloch scholarship: the exploitation of narrative and archive, of maps and place names, of ancient tools, engravings, imagery carved into bas-relief on buildings, even the earliest aerial photography, to recreate and understand the history of medieval Europe.

Fifty years later Bloch's insights belong almost in the category of truism, so completely have the humanistic sciences been transformed by interdisciplinarity and by a boundlessly catholic definition of documentation. Nearly everything is data now, not only for historians, but for archaeologists, classicists,

ecologists, ethno-botanists, anthropologists, sociologists, linguists, musicologists, and literary scholars. Medievalists, who once enjoyed and bemoaned the relative scarcity of textual documentation from the twelfth, thirteenth and fourteenth centuries, have learned to coax evidence from architecture and aerial photography. The shape of fields testifies about patterns of landholding and the evolution of agricultural technologies. European sinologists, cataloging the contents of a cave discovered in extreme Western China in 1930, were first intrigued by the text of Buddhist sutras written on scraps of silk. But far more interesting than the sutras were the household accounts for a monastic community – wage and price data and dietary information – scribbled on the verso of the silk fragments, as scarce writing material was recycled for a second use. The history of European family life has been deduced in significant measure from the examination of early paintings, drawings and etchings. Archaeologists remind us that the dirt surrounding excavated objects is as much *data* as the object itself. Climatologists and meterologists object strongly to the discard of LandSat images in which the earth's surface is obscured by cloud cover: the clouds are their *data*, even as they obscure the data of interest to geologists and ecologists.

From 1986 to 1991 RLG conducted three benchmark surveys of information needs in the humanities, social sciences and natural sciences, published in 1988, 1989 and 1991 respectively. These documents, again and again, the range of resources actively exploited by practitioners in various disciplines, driven by interdisciplinarity, by a pervasive interest in context and relationships, by the rise of feminist studies, and by a very broad definition of what is culture. We see that an anthropological definition of culture has come to dominate the historical sciences, raising interest in pulp magazines, folklore, advertisements, dissident literature and oral histories. We see visual materials – etchings, woodcuts, engravings and photographs – playing a growing role, not just in art history, but in historical research and cultural sociology. We are told that literary scholars are making increased use of images, and *will benefit* from automated cataloging of visual materials. We note that musicologists *have come to rely increasingly on unpublished materials as their interest has grown in the historical circumstances surrounding the creative process.* We see anthropology defined as a field science where primary data none the less includes tax rolls, minutes of village council meetings, geneologies, maps, photographs, and material culture artifacts. And in interview after interview, we are reminded that the institutions responsible for the preservation of and access to the documentary record of humankind have done a lot better job with secondary and narrative data than they have managed with archival materials, images, artifacts, and non-traditional data.

Almost from the beginning, RLG has worked to improve access to primary sources. In 1980 work was begun to define a standard descriptive format for archival and manuscript collections. In 1983 this work resulted in the so-called AMC (*Archival and Manuscript Control*) format. AMC is a MARC for-

mat, specially adapted to the description of primary, unpublished materials. The early adopters of AMC were university archivists, but in 1984 the format began to be used outside universities, especially by state archives and historical societies. Within ten years, dozens of repositories had created a union database of more than 250 000 AMC records, and by 1998, the 500 000th AMC record was entered into the RLIN database.

AMC is now widely accepted as an appropriate standard for collection-level descriptions of unpublished materials. As early as 1993, however, it became clear that AMC records could not provide adequate access to large collections of material, and that the adoption of AMC for series, folder, or item level description was not cost-effective. A project was therefore launched at the University of California, Berkeley, to enumerate and analyze the common elements in finding aids. This work led, in 1995, to the release of an alpha version of EAD (Encoded Archival Description) – a SGML-based document-type-definition for finding aids. RLG then played a key role in encouraging the adoption of EAD, including a program of training seminars given for members and non-members, in the United States, the UK and Australia. Other organizations assumed responsibility for the maintenance of EAD. The Library of Congress agreed to maintain documentation for the EAD standard, and the Society of American Archivists agreed to take on intellectual responsibility for its evolution and enhancement. In 1998, RLG entered into a contract with Apex Data Services to provide EAD markup services for archival finding aids. Apex has now established good staff expertise with EAD, and offers preferential rates to RLG members desiring to outsource conversion.

As the profile of EAD has been raised, a body of finding aids has been converted. Consequently, in September 1998, RLG was able to launch a new *Archival Resources* service, providing a single point of entry to EAD-encoded finding aids across the USA and around the world. Archival Resources now provides access to more than 4 000 finding aids describing collections in twenty institutions. A central index is maintained at RLG using OpenText software; the encoded finding aids are located on servers at the owning institutions. (As a temporary expedient, HTML copies of the encoded finding aids are also stored at RLG. This minimizes Internet-related response time delays for users, and eliminates the need for users to load SGML plug-ins into their browsers. When XML replaces HTML as an Internet standard and all browsers are naturally XML-aware, the need for this expedient will be eliminated.) The advanced searching capabilities of Archival Resources are rooted in the EAD markup on which it is based. However, RLG is now experimenting with the provision of more limited access to finding aids that have been digitized without benefit of EAD markup.

AMC, EAD and the Archival Resources service are part of a broad effort to improve access to primary materials. The core of this effort is ensure that descriptive metadata are created, eventually, for the entire record of man and his works, and that services are then created to navigate through the metadata to

target information. The published book, serial, map and sound recording collections of libraries are now reasonably well served. Thanks to AMC, EAD and related developments, collections of primary textual information have now also entered the managed information mainstream. Much more remains to be done with non-text objects of all kinds (artifacts), with geo-referenced information and with compiled statistical data.

Slides illustrating the Archival Resources service follow. The first three slides show the entrance to Archival Resources through RLG's Eureka search service, subsequent slides illustrate sample searches. A simple search for the names *Steiglitz* and *O'Keefe* is followed by an *advanced* search for *Roosevelt* as a personal name, *Alaska* as a geographic name, *expedition* as a keyword in an introduction, and *Harriman* as a keyword in the container list. This example demonstrates the powerful precision of information retrieval based on EAD markup.

Collaboratories – auf dem Weg zu neuen Formen der technisch unterstützten Kooperation

Von Elmar Mittler

Es gehört zu den vielen Vorurteilen der 60er Jahre gegen die Datenverarbeitung, daß sie den Menschen vereinsamt. Wir wissen heute, daß in vieler Hinsicht genau das Gegenteil der Fall ist. Die Datenverarbeitung führt zusammen. Schon bei der Einführung der bibliothekarischen Verbundsysteme konnte man feststellen, daß ausgerechnet die Katalogisierungsabteilung, die vorher in jeder Bibliothek gern *splendid isolation* pflegte, plötzlich großes Interesse daran hatte, Kollegen in anderen Bibliotheken kennen zu lernen, mit denen man über den Computer durch die gemeinsame Katalogisierung täglich zu tun hatte.

Neue Formen der Kooperation

So wie die Bibliothekare näher zusammengerückt sind, hat die Digitalisierung auch dazu geführt, daß Bibliothekare, Computerspezialisten und Wissenschaftler in einer Form eng zusammenarbeiten, die wir bisher vergleichbar nicht kannten. Neuartige Teams und gemeinsame Service- und Organisationsformen sind erforderlich, um die technischen Voraussetzungen für funktionierende digitale Bibliotheken zu schaffen.

Was die Bibliothekare erlebt haben, erleben die Wissenschaftler selbst in ähnlicher Weise. In der Vergangenheit waren am ehesten die Naturwissenschaftler, z.B. die Atomphysiker oder Astronomen, daran gewöhnt, weltweit zusammenzuarbeiten, weil die Forschungsinstrumente in internationaler Kooperation betrieben wurden. Nicht ohne Grund ist die Software für das WWW letztlich in Ispra entwickelt worden, einem internationalen Großforschungszentrum der Physiker in Europa.

E-Mail und WWW haben aber inzwischen auch die Arbeitsbedingungen der Geisteswissenschaftler wesentlich verändert. Zwar gibt es auch jetzt noch den Typus des gelehrten Einzelgängers, der seine große Monographie zu Hause mit Tinte und Feder schreibt. Aber er wird zunehmend zur Ausnahme. Das muß nicht unbedingt ein Fortschritt sein. Aber es öffnen sich neue Möglichkeiten, die Arbeiten realisierbar machen, die es bisher in dieser Weise noch nicht gegeben hat.

Wichtiges Ausgangsglied der neuen Aktivitäten sind die Diskussionsgruppen und Mailinglisten, die es in jedem Fachgebiet gibt. Sie sind inzwischen so vielfältig geworden, daß es längst Listen über Diskussionslisten gibt. Ähnlich

ist die Entwicklung der elektronischen Zeitschriften, der e-journals: sie haben Zuwachsraten an neuen Titeln, die weit über denen der gedruckten Zeitschriften liegen.

Noch entscheidender aber ist die Entwicklung von Editionen oder ganzen Text-Corpora in verschiedenen Fachgebieten geworden, der Anglistik ebenso wie der Geschichte oder der Theologie. In ihrer einfachsten Form bieten sie Bilder von Handschriften teilweise kombiniert mit den Texten; der nächste Schritt ist, sie auch mit Annotationen zu versehen; im Idealfall sind es Hypertext-Corpora, in denen die Texte mit Kommentaren und der zugehörigen Literatur verbunden sind.

Es ist die Chance des Internets, derartige Werke arbeitsteilig in collaboratories zu entwickeln und verteilt zu speichern. Hierfür können Beiträge von den unterschiedlichsten Seiten eingebracht werden, auch wenn viele Corpora von führenden Forschungszentren moderiert werden. Diese Zentren sind oft in den Vereinigten Staaten – wie das *Center for Electronic Texts in the Humanities* (Princeton and Rutgers Universities) *(http://www.ceth.rutgers.edu)* – oder in England – wie das *Oxford Text Archive (http://ota.ahds.ac.uk)* – konzentriert. Das *International Computer Archive of Modern and Medieval English* ICAME *(http://www.hd.uib.no/icame.html)* aber liegt z.B. in Bergen in Norwegen.

Was hier in wachsender Zahl und großem Umfang allgemein für ganze Forschungsgebiete oder spezialisiert auf einzelne Autoren, Zeiten oder Regionen entsteht, ist eine völlig neue Forschungsumwelt. Die Aktivitäten der Archive und Bibliotheken gliedern sich hier ein. Das Beste ist es, wenn sie diese aktiv mitgestalten, wie dies mit der Verlagsgründung Highwire z.B. der Fall ist, die neue Formen verlinkten Publizierens entwickelt hat. Es gibt viele andere Beispiele, vor allem in den USA. Von der erfolgreichen Arbeit der Digital Library der Columbia University *(http://www.columbia.edu/cu/librairies/digital)* in New York konnte z.B. Carol Mandel berichten. Entscheidende Herausforderung der Zukunft ist es – Deanna Marcum hat es deutlich gemacht – die einzelnen digitalen Bibliotheken der Welt zu vernetzen. Dies kann nicht von oben nach unten – top down – befohlen werden; man muß die bestehenden und sich entwickelnden Einrichtungen bottom up zusammenführen.

Dazu sind drei Aspekte wichtig:[1]

- technische Interoperabilität: Hardware, Netzwerke, Datentypen, Protokolle.
- Informations-Interoperabilität: Metadaten, Authority files und user interfaces.
- soziale Interoperabilität: Aufstellen von Spielregeln, Definition von Rechten und Pflichten von Personen und Organisationen.

---

[1] Vgl.: Edward A. *Fox* and Gary *Marchionini:* Toward a worldwide digital library. In: Communications of the ACM 41 (1999) 4 S. 29–32, hier S. 30.

Der Beitrag der Bibliotheken

Im Bereich der Protokolle ist es ein wichtiger Fortschritt, daß es den Bibliothekaren gelungen ist, mit der Z 39.50-Schnittstelle einen Standard zu schaffen, der heute ohne größere Schwierigkeiten den gemeinsamen Zugriff auf die verschiedensten Kataloge und Bibliographien ermöglicht. Inter-Library-Loan und Document-Delivery-Systeme können darauf aufgesetzt werden. Problematisch wird die praktische Umsetzung international allerdings, wenn es um die Bezahlung der entstehenden Kosten geht. Im Rahmen des *German Resources Project (http://lcweb.loc.gov/loc/german)*, eines gemeinsamen Projektes der Association of Research Libraries und deutscher Sondersammelgebietsbibliotheken, werden deshalb für diese spezielle Frage Modelle entwickelt und erprobt. Dabei sind in der Regel gedruckte Materialien die Ausgangsbasis, die aber mit Hilfe der Software ARIEL *(http://www.rlg.org/ariel.html)* auch als digitales Material versandt werden können.

Schwieriger ist es, digitales Material zu erschließen. Im Beitrag von Lossau ist deutlich geworden, wie im *American Memory Project* unterschiedliche Ressourcen verschiedener Bibliotheken zusammengeführt und erschlossen worden sind. Dabei ist es wichtig, daß nicht nur der Titel einer digitalen Sammlung, sondern die Einzelstücke und die Inhalte möglichst detailliert zugänglich werden. Tools sind auch dafür teilweise entwickelt. Es kommt jetzt darauf an, sie zu standardisieren und gemeinsam einzusetzen. Das ist z.B. ein Grund dafür, daß die europäischen Bibliotheken, die sich im Projekt DIEPER (Digitized Electronic Periodicals) *(http://www.sub.uni-goettingen.de/gdz/dieper)* zusammengeschlossen haben, um Zeitschriften gemeinsam zu digitalisieren und inhaltlich zugänglich zu machen. Sie suchen den Kontakt und die Zusammenarbeit mit JSTOR *(http://www.jstor.org)*, das an der Zeitschriftendigitalisierung in den Vereinigten Staaten arbeitet.

Die allgemeine Einführung von Standards

Für die Erschließung ist auch die internationale Metadatendiskussion *(http://www.sub.uni-goettingen.de/f_projek.htm* Unterpunkt *Metadaten-Server)* von Wichtigkeit. Carol Mandel hat das Modell des Master-Metadata-Files der Columbia University gezeigt. Ideal wäre es natürlich, wenn die dort gespeicherten Daten in standardisierter Form vorliegen könnten. Dublin-Core ist hierfür ein neuer Ansatz, der auch dann ein positives Ergebnis gebracht hat, wenn seine Anwendungen sich nun zunehmend zu spezifizieren und zu zersplittern scheinen. Denn es ist durch Dublin-Core zum ersten Mal gelungen, daß Bibliotheken, Archive und Museen, aber auch Fachgesellschaften und Wissenschaftlergruppen, nicht zuletzt aber auch die Entwickler des WWW, von einem Set an Grundelementen bei der Beschreibung ausgehen. Mit dem

WWW gibt es die einmalige Chance, die unterschiedlichsten Einrichtungen virtuell zusammenzuführen, die bisher oft völlig getrennt, sozusagen in verschiedenen Welten, ihre Daten erschlossen haben. Die Forscher aber haben wenig Verständnis dafür, warum z.B. eine mittelalterliche Handschrift, die in einer Bibliothek, einem Archiv oder in einem Museum aufbewahrt wird, deshalb je nach dem mehr oder weniger zufälligem Standort in völlig verschiedener, nicht kompatibler Form erschlossen wird.

Ein Beispiel, wie in konsequenter Weise die Infrastruktur für die Kommunikation auf einheitlicher Basis aufgebaut werden kann, ist das vom Deutschen Forschungsnetz geförderte MathNet-Programm *(http://www.mathnet.de)*, das den einheitlichen Inhalt von Webseiten der mathematischen Institute in Deutschland ebenso definiert, wie es Vorgaben für das elektronische Publizieren gibt. Die Verbindung zu den Bibliotheken ist, im von der DFG geförderten *Math-Net-Bib-Projekt*, ergänzend in Entwicklung. Eine Übertragung der so geschaffenen flexiblen Struktur auf die europäischen und die internationalen mathematischen Fachgesellschaften ist in Vorbereitung.

Der nächste wichtige Schritt, den Bibliotheken gemeinsam bewältigen müssen, ist den Zugang zu den Materialien zu erleichtern. Hier gibt es technische Probleme vielfältiger Art, nicht zuletzt das Problem der dauerhaften Sicherung und des schnellen Zugriffs. Auch hier braucht man Standards und stabile Kooperationen zwischen Bibliotheken. Für viel gebrauchtes Material und zur Sicherung wird dabei immer wieder auch das *mirroring*, das Spiegeln von digitalen Materialien, nicht allein das Verlinken sinnvoll sein.

Die virtuelle Fachbibliothek

Es ist dabei sicher ein praxisnaher Ansatz, fachliche Schwerpunkte zu bilden, weil die Interessen und Methoden der Informationsversorgung in verschiedenen Fachgebieten stark divergieren. Die Deutsche Forschungsgemeinschaft ist dabei, die Entwicklung virtueller Fachbibliotheken zu fördern, die auf das erfolgreiche System des Sondersammelgebietsprogramms aufsetzt. Als Sondersammelgebietsbibliothek für Angloamerikanische Kultur und Geschichte hat z.B. die Niedersächsische Staats- und Universitätsbibliothek Göttingen engen Kontakt mit Wissenschaftlern, die sich für die gegenseitige Beeinflussung der amerikanischen und der deutschen Geschichte interessieren. Sie haben eine Liste zusammengestellt, der man entnehmen kann, was Forscher benötigen, um über den Atlantik hinweg in collaboratories gemeinsam arbeiten zu können: Sie wünschen sich

1. eine bessere Verzeichnung der archivarischen Quellen z.B. zur Geschichte der Einwanderung oder bestimmter Religionsgemeinschaften, auf beiden Seiten des Atlantiks, in Archiven und Bibliotheken;

2. eine digitale Kernbibliothek zur amerikanischen und deutschen Geschichte, die einem unabhängig vom Standort das grundlegende Arbeiten an Fragen der amerikanischen Geschichte ermöglicht;
3. gezielte Programme für die Digitalisierung von Materialien, für die es konkrete Forschungsvorhaben gibt, die nicht nur punktuell sein dürfen, sondern zum Teil als langfristige Forschungsprojekte angelegt sein sollten.

Bibliotheken, Archive und Museen sind aufgerufen, dafür gemeinsam die Grundlagen zu schaffen, das Environment für die Collaboratories auch der geisteswissenschaftlichen Forschung der Zukunft. Es ist positiv zu bewerten, daß die Europäische Kommission gemeinsame Aktivitäten dieser Gruppen im nächsten Förderprogramm mit Vorrang unterstützen will.

Ansätze zur amerikanisch-deutschen Zusammenarbeit

Der Weg dazu kann nur über eine Intensivierung der internationalen Zusammenarbeit zwischen den verschiedensten Bibliotheken und auf den verschiedensten Ebenen sein. Es ist deshalb ein besonderer Erfolg, daß der Antrag auf Unterstützung, den die SUB Göttingen an die Mellon-Foundation gestellt hat, positiv beschieden wurde. Damit ist es möglich, im Rahmen des German Resources Project der Association of Research Libraries durch gegenseitigen Austausch Lösungen auch für praktische Fragen wie Dokumentenlieferung, Angleichung der Katalogregeln, der Zusammenarbeit bei der Erschließung von Web-Seiten und der Katalogisierung des Internet voranzukommen. Gemeinsame Sitzungen, die in Washington (Juli 1998) und Göttingen (März 1999) stattfanden *(http://lcweb.loc.gov/loc/german/minutes.html)*, lassen schon erste praktische Erfolge sichtbar werden.

Ebenso wichtig aber ist es, daß auf der Ebene der großen Initiativen, wie Commission of Information and Library Resources, Randbedingungen für die Standardisierung der digitalen Bibliothek und Projekte für ihre praktische Umsetzung vorbereitet werden. Im Bereich der Lizenzen ist aus der Erkenntnis, daß die Verleger international arbeiten, aus der Consortium of Consortia ICOLC die International Coalition of Library Consortia *(http://www.library.yale.edu/consortia)* geworden. Vielleicht stehen wir am Anfang einer Entwicklung, die aus der Digital Library Federation *(http://www.clir.org/diglib/dlfhomepage. htm)* eines Tages auch eine IDLF, eine International Digital Library Federation werden läßt. Auch Fördereinrichtungen wie die National Science Foundation und die Deutsche Forschungsgemeinschaft haben diese Notwendigkeit erkannt und Programme zur Förderung gerade auch der internationalen Zusammenarbeit im Bereich der digitalen Bibliotheken ausgeschrieben.

Ein regelmäßiger Erfahrungsaustausch der Mitglieder der Digital Library Federation und der Praktiker der Digitalisierungsprojekte in Deutsch-

land, organisiert von den Digitalisierungszentren Göttingen *(http://www.sub.uni-goettingen.de/gdz)* und München *(http://www.bsb.badw-muenchen.de/mdz)*, wäre sicher ein ergänzender guter Weg, erfolgreich in der Zusammenarbeit voranzukommen. Ein Meilenstein amerikanisch-deutscher Kooperation im Bereich der digitalen Bibliothek war sicher das von der Landesarchivdirektion Baden-Württemberg veranstaltete Ludwigsburger Kolloquium im November 1998.

Organisation und Technik der Digitalisierung von Archiv- und Bibliotheksgut

Planung und Durchführung von Digitalisierungsprojekten

Von Marianne Dörr

Für ihre traditionellen Aufgaben des Sammelns, der Erschließung, des Nachweises und der dauerhaften Bereitstellung von Materialien haben Bibliotheken und Archive Arbeitsverfahren, Geschäftsgänge und Regelwerke etabliert, die eine einigermaßen ökonomische und gleichmäßige Abwicklung erlauben. Ähnliche Handreichungen wünschen sie sich für die neue Aufgabe der Digitalisierung. In den USA, wo längere Erfahrungen mit der Digitalisierung von Sammlungen vorliegen, existieren auch bereits umfangreiche Digitalisierungs-*Fahrpläne* wie z.B. die Checkliste aus dem *National Digital Library Program* der LoC.[1]

Zu neun großen Paketen gruppiert – von der Auswahl der Sammlung bis zum Update der fertigen Anwendung – werden hier die einzelnen Aspekte und Arbeitsschritte von Digitalisierungsprojekten zum Abhaken aufgeführt. Für die Erstellung funktionierender Checklisten gibt es jedoch Voraussetzungen:

- einen Grundkonsens über die Zusammenstellung von digitalen Sammlungen,
- eine prinzipielle Einigung über die Präsentation, die Tiefe der Erschließung und die daraus resultierenden Zugriffsformen,
- das Wissen über das technische System, die technische Realisierung der Bereitstellung und vor allem:
- Erfahrung, die aus der Abwicklung einer kritischen Menge an *homogenisierbarem* Material resultiert.

Im Vergleich mit den USA liegen in Deutschland erst wenige Erfahrungen mit der Gesamtabwicklung großer Projekte vor. Die vielfältigen inhaltlichen, organisatorischen und technischen Rahmenbedingungen für Digitalisierungsprojekte sind oft noch nicht präzise abgesteckt. Deshalb möchte ich keine derart detaillierte Checkliste aufstellen, sondern auf wenige grundsätzliche Fragestellungen eingehen, die von allen Institutionen bei der Planung von Digitalisierungsprojekten zu bedenken sind. Dabei sollen die bereits verfügbaren

---

[1] Vgl. URL: *http://memory.loc.gov/ammem/ftpfiles.htm.*

konkreten Erfahrungen aus eigenen bzw. aus den dokumentierten ausländischen Projekten einfließen.[2]

Auf folgende Themen werde ich näher eingehen:

1. die Auswahl von Projekten und die inhaltliche Projektkonzeption
2. das Erschließungskonzept
3. Arbeitsschritte der Durchführung: Aufwands- und Personalabschätzungen, Eigenarbeit und Outsourcing
4. Bereitstellung/technische Lösung
5. Ergebnisbewertung

Diese Schritte sind nicht als sequentielle Abfolge zu verstehen, sondern in Realität eng miteinander verwoben.

Die Auswahl von Projekten oder *was wird digitalisiert?*

Die amerikanische Checkliste beginnt mit der lapidaren Aufforderung: *Select a collection for digital conversion!* Doch bereits die Selektion von Materialien für Digitalisierungsprojekte ist eine komplexe Aufgabe. Digitalisierungsprojekte stellen gegenüber den oben angesprochenen traditionellen Aufgaben von Bibliotheken und Archiven etwas Neues dar: Mit Digitalisierungsprojekten soll nicht primär etwas nachgewiesen, verwaltet und für eine Benutzung – bis in ferne Zukunft – aufbewahrt und gesichert werden, sondern Digitalisierungsprojekte sind als eine Form von *electronic publishing* anzusehen. Inhalte werden für ein Publikum aufbereitet, aktiv nach außen gebracht. Aber was aus einem oft riesigen Bestand soll ausgewählt werden? Auch bei der vorläufig noch vorherrschenden Beschränkung auf urheberrechtsfreie Materialien bleibt die Auswahl immens.

Die Zielvision einer *Verteilten digitalen Forschungsbibliothek*, die dem Förderprogramm der Deutschen Forschungsgemeinschaft seinen Namen gab, gibt nur einen sehr weitgefaßten Anhaltsrahmen, aus dem sich derzeit Projekte verschiedener Gattungen, Inhalte und unterschiedlicher Motivationen wie Inseln hervorheben (von Turfanhandschriften zu DDR-Untergrund-Zeitschriften). Die Möglichkeit, ein Korpus aufzubauen, das sich alleine trägt, wie es die Bibliothèque nationale de France mit der Digitalisierung von 86 000 Werken getan hat, steht in Deutschland wohl keiner Institution offen.

Bei Projektplanungen sollten deshalb früh die strategischen Überlegungen zur Situierung des eigenen Projekts reflektiert werden. Einige wichtige Kriterien, die die Auswahl mitbestimmen können:

---

[2] Vgl. zu weiteren Materialien, die Literaturliste im Anhang, die als Einstiegshilfe dienen soll.

- Wie ist die derzeitige Nutzung, das Forschungsinteresse?
- Gibt es bereits inhaltlich verwandte Projekte? (zum einen zur Überprüfung der dortigen Nutzungsintensität, zum anderen auch, um bewußt an existierende Projekte und Sammlungen anzuknüpfen. Das Internet bietet das Forum für eine weltweite Recherche)
- Gibt es Verlagsaktivitäten in dem Bereich?
- Sind über die digitale Aufbereitung neue Nutzungsformen und Erkenntnisse möglich?
- Eignet sich die Sammlung als Kernbestand für eine Ausweitung?

Ein weiteres legitimes Motiv kann darin liegen, einem breiteren Publikum in einer Art virtueller Ausstellung Zugang zu einem seltenen und/oder wertvollen Bestand zu geben (und damit implizit auch Öffentlichkeitsarbeit für die eigene Institution zu betreiben).

Nach Möglichkeit sollte von Anfang an der Einbezug von anvisierten Nutzergruppen mit ihren Fragestellungen und Bedürfnissen angestrebt werden – mindestens über eine kritische Prüfung von Projektvorschlägen durch Vertreter der anvisierten Zielgruppen. Ideal sind Kooperationen schon bei der Antragstellung, z.B. durch Vereinbarung von konkreten Nutzungen. So könnten digitalisierte Materialien Gegenstand eines Seminars werden, ein Teil der (fachlichen) Erschließungsarbeit von den studentischen Nutzern eingebracht werden. Auch eine Zusammenarbeit mit Schulen bzw. den Fachleitern an Studienseminaren könnte erprobt werden – eine Mehrfachnutzung der digitalisierten Bestände wäre in diesem Fall eher möglich als bei der hohen Spezialisierung an Universitäten. Dieses Verständnis von Digitalisierungsprojekten setzt bei Bibliotheken und Archiven ein aktives Nach-außen-gehen, Bereitschaft zum Dialog und zur kritischen Revision eigener Vorstellungen über das Interesse und die Bedürfnisse der Nutzer voraus. Vorbildlich sind viele Projekte im Bereich der amerikanischen Universitätsbibliotheken, die in sehr engem Kontakt mit der Lehre arbeiten. Naturgemäß ist die Zielgruppe bei Staatsbibliotheken und Archiven schwieriger zu bestimmen. Digitalisierungsprojekte könnten bei diesen Institutionen Anlaß zur Standortbestimmung werden. Das Ziel ist klar zu definieren: die Vermeidung von Fehlinvestitionen und *digitalen Friedhöfen*, auf die sich allenfalls eine Suchmaschine einmal verirrt.

Erschließung

Von der inhaltlichen Projektkonzeption ist die Frage der Erschließung und des Angebots an Zugriffspunkten nicht zu trennen. Die Beschreibung der Zugriffsszenarien, die dem Benutzer angeboten werden sollen – und die Frage nach der technischen und der arbeitsökonomischen Realisierbarkeit – gehören zu den wichtigsten Arbeitsschritten bei Digitalisierungsprojekten überhaupt.

Die Erwartungshaltungen des Publikums schwanken im elektronischen Bereich noch stark: Einerseits herrscht Freude darüber, überhaupt Material zu einem Thema im Web zu finden – und sei es nur als Imagedaten zum Ausdrucken. Andererseits besteht ein Anspruch, der manchmal die Realität des Leistbaren fern hinter sich läßt, z.B. daß im Zuge der Digitalisierung alle Schwächen der zugrundeliegenden Publikationen gleich mitbeseitigt werden könnten, neue Editionen, neue tief erschlossene Kompilationen entstünden. Oder man hofft auf Suchtools, die tatsächlich neue Möglichkeiten der Benutzung, der Analyse und damit der Erkenntnis bringen. Sicherlich werden wir zunehmend mit der Erwartung nach multimedialer Aufbereitung konfrontiert. Bibliotheken und Archive müssen für ihre Projekte realistische, d.h. auch finanzierbare Zwischenwege finden.

Auch bei der Erschließungskonzeption sollten eine Reihe von Fragen vorab reflektiert werden:

- Welche Fragen hat der Benutzer an das Material?
- Welche Zugriffspunkte müssen deshalb realisiert werden?
- Welche Beschreibungsdaten existieren bereits und können übernommen werden?
- Was muß neu erstellt werden, wie groß ist der Aufwand einzuschätzen?
- Gibt es bei vergleichbaren Materialien Standards, die sich etabliert haben und nicht unterschritten werden sollten?
- In welcher Struktur, in welchem Format sollten die Daten erfaßt werden?
- Ist intellektuelle Erschließung notwendig?
- Kann nach Vorgabe von Hilfskräften erfaßt werden?
- Stehen Möglichkeiten der automatischen Indexierung zur Verfügung?
- Eignet sich das Material für eine automatische Schrifterkennung?

Beim zweiten Kolloquium der beiden DFG-geförderten Digitalisierungszentren Göttingen und München, das im Oktober 1998 in München stattfand, stand die Frage der Erschließung im Zentrum der Veranstaltung.[3] Dabei wurde auch deutlich, daß zwei Stränge zu trennen sind:

Der eine Strang betrifft die inhaltlichen/fachwissenschaftlichen Erschließungsfragen, z.B. ob und welche Systematik zur Klassifikation der Materialien und zum Zugriff eingesetzt wird, ob nach einem Thesaurus beschlagwortet wird etc. Das sind die klassischen Probleme der Fachwissenschaft, der inhaltlichen Sacherschließung, der Fachdokumentation.

---

[3] Das Kolloquiumsprogramm ist zugänglich unter der URL: *http://www.bsb.badw-muenchen.delmdz/koll.htm.*

Der andere Strang betrifft die *Verpackung* der Inhalte in Strukturen und Formaten: Hier ist die Frage zu stellen, ob Codierungsstandards, Dokumentbeschreibungssprachen eingesetzt werden können und ob dafür Vorlagen existieren. Wir bewegen uns auf der Ebene der Austauschformate.
In amerikanischen Projekten haben sich für das Angebot von Texten und historischen Quellensammlungen bestimmte Standards etabliert: relativ flache Strukturen, d.h. einfache (nicht zu tief gehende) SGML-Codierungen, deren Umsetzung einen Zugriff über Browsing (Titellisten, dann Aufblättern der Struktur der Einzeldokumente) und einen Suchzugriff auf der Ebene von Dokumenttiteln bzw. Dokumentuntergliederungen ermöglicht. Bei vorher gedruckten, textuellen Dokumenten ist in der Regel der Volltext vorhanden und über Volltextsuchmaschinen zugreifbar. Ein reines Image-Angebot mit Zugriffsmetadaten wird sich vermutlich langfristig nur bei nicht rein textbasiertem Material behaupten können. Als Beispiel für übertragbare Codierungsvorgaben für diverse Text- und Dokumententypen möchte ich für den bibliothekarischen Textbereich die Arbeiten der *Text Encoding Initiative*, für die Archive die *Encoded Archival Description*[4] nennen, die von der LoC und der Society of American Archivists gepflegt wird.

Dieser zweite Strang ist als neues Aufgabengebiet der Bibliothekare und Archivare, die sich mit Digitalisierung befassen, zu sehen. Auf dem Gebiet der neutralen Standards, der *Verpackungen* muß Know-how aufgebaut werden, um die Daten austauschbar und *haltbar* zu machen. Aber auch die Einarbeitung in und die Anwendung dieser Beschreibungssprachen erfordert Zeit- und Ressourcen-Einsatz.

Ein dritter Bereich, die zusätzliche Entwicklung von Software-Modulen zur Verbesserung des Zugriff und der Arbeit mit dem Material ist teuer. Aus dem philologischen Bereich sind Tools zur morphologischen Analyse bekannt. Im amerikanischen humanistischen PERSEUS[5]-Projekt wurde so ein Tool für die Analyse griechischer Texte entwickelt. Anregungen hierzu müssen in jedem Fall aus der Fachwissenschaft kommen, Bibliotheken und Archive werden wenig initiativ sein können. Im verwandten Förderschwerpunkt der Deutschen Forschungsgemeinschaft auf dem Gebiet der Informatik *Verteilte Vermittlung und Verarbeitung Digitaler Dokumente* (kurz V3D2) sind kooperative Software-Entwicklungen im Gange. Die Aufgaben der Bibliotheken und Archive werden jedoch vor allem auf den Gebieten der Nutzung von und der Mitarbeit bei Standardisierungen, Nachnutzung von Entwicklungen, weniger im Bereich der Neukonzeptionen liegen können.

---

[4] Vgl. die URLs im Anhang.
[5] Vgl. URL: *http://www.perseus.tufts.edu.*

Arbeitsschritte der Durchführung

Welche Arbeitsschritte fallen nach der inhaltlichen Entscheidung und dem Entwurf des Erschließungskonzepts in der konkreten Umsetzung von Projekten an. Ohne Anspruch auf lückenlose Vollständigkeit sollen folgende Tätigkeiten genannt werden – einige von ihnen könnten durchaus Gegenstand eigener Vorträge sein.

- Zusammenstellung des zu digitalisierenden Materials
- Materialanalyse- und prüfung
- Vorentscheidung über die Art der Digitalisierung
- Zusammenstellen von Testsamples für eine Digitalisierung
- Festschreibung der technischen Parameter und der Anforderungen für unterschiedliche Bildformate (z.B. Archiv- und Netzversion)
- Vereinbarung der Modalitäten der Datenlieferung (Datenträger…)

In anderen Beiträgen zu diesem Kolloquium war viel von den technischen Standards der Digitalisierung, den Möglichkeiten der Digitalisierung vom Film und direkt vom Original, dem Problem des Farbmanagements etc. die Rede. Das soll an dieser Stelle nicht wiederholt werden. Nach einer Materialanalyse und unter Berücksichtigung weiterer Aspekte (Bestandserhaltung) muß die Entscheidung über die technische Durchführung getroffen werden. Die Digitalisierung i.S. von reinem Scannen kann outgesourct werden – der Markt an Dienstleistern ist zwar noch nicht so groß wie er bei den Mikroverfilmern war, wächst aber noch. Schwieriger ist die Einschätzung der organisatorischen Arbeiten, die nicht delegiert werden können.

- Einholen von Angeboten, Vergleich, Auswahl des Dienstleisters
- Festlegung von Verzeichnisstrukturen und Dateinamen
- Aufstellung von Konkordanzen zwischen Imagefolgen und vorgegebenen Strukturen (Seitenzählungen, den Angaben unter denen man etwas wiederfinden will)
- Erstellung von Metadaten für bibliographische, inhaltliche, technische Beschreibung bzw. Verfahren der Übernahme aus vorhandenen Systemen

Bei diesen Arbeitsschritten wird jedes Projekt einen unterschiedlichen Aufwand erfordern. Ein Beispiel aus den Projekten der Bayerischen Staatsbibliothek: Bei den Reichstagsberichten aus dem 19. Jahrhundert mit zusammen über 110 000 Seiten mußten allein für die Erstellung von Excel-Listen zur Dokumentation (und zum späteren automatisierten Import ins Bereitstellungssystem) der Bandstrukturen (Beginn der Sitzungen etc.) und Konkordanz zwischen Imagefolge und Seitenzählung im gedruckten Buch 160 Stunden an Hilfskraft-Arbeit investiert werden.

Auch die Metadaten-Erstellung für die Erschließung muß im Regelfall selbst erledigt werden. Für Standardmaterialien können Arbeiten wie die SGML-Codierung nach vorgegebenen Dokumenttypdefinitionen zunehmend outgesourct werden. Aber es müssen klare Vorgaben erstellt und ihre Einhaltung überwacht werden. Generelle Regeln sind nicht möglich. Die Durchführung eines Projekts mit einer großen Menge an homogeneren Daten (z.B. Zeitschriftenjahrgänge mit Standard-Erschließung auf Artikelebene) sind einfacher abzuhandeln als die Aufbereitung einer kleineren Menge an heterogenem Material, z.B. ein Nachlaß aus Bildern, handschriftlichem und gedruckten Materialien unterschiedlicher Größe, in unterschiedlichem Erhaltungszustand mit nur rudimentär vorhandenen Erschließungsdaten. Auch wenn diese Projekte oft reizvoller sind und mehr Erfolg bei einer Antragsstellung in Aussicht stellen, ist doch dringend zu raten mit kleineren und einfacheren Materialien zu beginnen.

- Erstellung eines Übergabeprotokolls und Entwicklung von Worksheets für oder mit dem Dienstleister mit Kontroll- und Reklamationsfeldern
- Entwicklung eines Verfahrens für die Qualitätskontrolle

Die Qualitätskontrolle stellt einen wesentlichen Arbeitsfaktor dar. Für sie muß – auch bei outgesourcter Digitalisierung – ein adäquater Arbeitsplatz in der jeweiligen Institution vorhanden sein. Da bei der Beurteilung von Bildqualitäten viele Faktoren eine Rolle spielen, sollte man sich hier mit dem Dienstleister auch über Bildschirmqualitäten abstimmen, evtl. Referenzbilder hinterlegen. Aus amerikanischen Projekten wird eine stichprobenartige Qualitätskontrolle bei ca. 10% der Bildfiles empfohlen – das kann bei Projekten mit mehr 100 000 Images in mehreren Formaten einen nicht zu vernachlässigenden Zeitfaktor ausmachen. Aus qualitätsvollen Farbdigitalisierungen können Dateigrößen des Einzelbildes von 50 MB und mehr resultieren, so daß für die Kalkulation der Qualitätskontrolle auch die Dauer der Bildladezeiten in Rechnung gestellt werden muß.

Die Basiserfahrung der bisherigen Projekte weist einen noch sehr hohen Anteil an Eigenarbeit auf. Wenn man die vom Dienstleister zu erbringende Scanzeit als 100% betrachtet, so sind zwischen 30% und 40% (bei Erstprojekten eher mehr) an Eigenanteil im Bereich Vorbereitung, Qualitätskontrolle, Nachbereitung zu erbringen. Dabei kann ein Großteil dieser Arbeiten von Hilfskräften erledigt werden, aber die Erstellung von Vorgaben und die Überwachung der Einhaltung erfordert weiterhin einen hohen Ressourcenverbrauch beim Stammpersonal. Mit erfahrenen und zuverlässigen Dienstleistern kann zunehmend die Übernahme von mehr Positionen vereinbart werden, bei den ersten Projekten lernen aber – wie die Erfahrung in den bisherigen Projekten gezeigt hat – meistens noch beide Seiten.

Bereitstellung / technische Lösung

Die letzten Schritte umfassen die Präsentation und die Verfahren zur Verwaltung und Sicherung der digitalisierten Daten.

- Datenimport ins eigene System: Speicherung, Verwaltung, Archivierung

Die Durchführung der Digitalisierungsschritte kann besser gestaltet werden, wenn die Art der Bereitstellung, das technische System bereits vorhanden und bekannt ist. Bei den gegenwärtigen Projekten wird allerdings oft noch mit mehreren Variablen gearbeitet, was eine effiziente Durchplanung bzw. ein Abstimmen der Schritte aufeinander erschwert. Mögliche Arten der Bereitstellung sind das Thema verschiedener Vorträge und Präsentationen des Kolloquiums, so daß ich nicht näher darauf eingehen will. Ich möchte nur darauf hinweisen, daß hier ein weiterer Arbeitsfaktor liegt, der unterschiedlich große Anforderungen stellen kann: Wenn die codierten Daten in einem Filesystem abgelegt und in HTML-Seiten mit rein navigierendem Zugriff angeboten werden, ist weniger zu tun als wenn eine Datenbankanwendung erstellt wird und differenzierte Definitionen der Felder, Feldlängen, Feldtypen und Suchzugriffe notwendig sind. Die Anschaffung und Installation eines Bereitstellungssystems, das bestimmte Investitionen in Hard- und Software erfordert, lohnt sich meist nicht für kleinere Institutionen mit nur gelegentlichen Projekten. Hier sind einerseits *kleine*, pragmatische Lösungen gefragt, andererseits aber die Kooperation mit größeren, zentralen Institutionen, die dadurch einen gewissen Dienstleistungsauftrag erfüllen.

Ergebnisbewertung

Dieser Punkt steht am Ende eines jeden Projektes. Bei der neuen Aufgabe der Digitalisierungsprojekte, in der Bibliotheken und Archive ihre Rolle und ihr Können neu zu bestimmen haben, ist er aber von besonderer Bedeutung. Eine kritische Dokumentation auch der Fehlschritte kann beim nächsten Projekt viel Zeit sparen. Statistiken über die Nutzungshäufigkeiten und Nutzerbefragungen sollten besonders eine Kontrolle der inhaltlichen Auswahl und der Erschließung ermöglichen. Allerdings ist – auch dies für öffentliche Institutionen noch ungewohnt – eine Art von Reklame, oder besser Öffentlichkeitsarbeit zu leisten.

Digitalisierungsprojekte sind eine neue Aufgabe der Bibliotheken und Archive, vor allem handelt es sich – und das wird oft unterschätzt – um eine sehr komplexe Aufgabe. Damit sind Fehler unvermeidbar und vermutlich muß eine Reihe von Fehlern gemacht werden, bis auch die deutschen Bibliotheken und Archive praktikable Checklisten für Digitalisierungsprojekte aufstellen können.

Literaturhinweise

Inhaltliche Projektkonzeption

Dan *Hazen,* Jeffrey *Horrell* and Jan *Merrill-Oldham:* Selecting Research Collections for Digitization. CLIR, August 1998.
http://www.clir.org/pubs/reports/hazen/pub74.htm

Stephen E. *Ostrow:* Digitizing Historical Pictorial Collections for the Internet. CLIR, ECPA 1998.

Retrospektive Digitalisierung von Bibliotheksbeständen. Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppen „Inhalt“ und „Technik“. (dbi-materialien 166). Berlin 1997.

Projektplanung

Carl *Fleischhauer:* Steps in the Digitization Process (1996).
http://lcweb2.loc.gov/ammem/award/docs/stepsdig.html

Carl *Fleischhauer*: Digital Historical Collections: Types, Elements, And Construction (1996).
http://memory.loc.gov/ammem/elements.html.

NDLP Project Planning Checklist.
http://lcweb2.loc.gov/ammem/prjplan.htm.

Ausschreibungsunterlagen

LoC Requests for Proposals (RFP), alle verfügbar unter:
http://memory.loc.gov/ammem/ftpfiles.html

RFP96-18: „Digital Images from Original Documents Text Conversion and SGML-Encoding“

RFP96–5: „Conversion of Microfilm to Digital Images“

RFP97–9: „Conversion of Pictorial Materials to Digital Images“.

RLG Model Request for Proposal (RFP) for Digital Imaging Services (1997).
http://lyra.rlg.org/preserv/RLGtools.html

RLG Worksheet for Estimating Digital Reformatting Costs (Modell zur Kostenermittlung)
http://lyra.rlg.org/preserv/RLGtools.html.

Codierung/Dokumentenbeschreibung

Text Encoding Initiative
http://www.uic.edu/orgs/tei/
Encoded Archival Description
http://lcweb.loc.gov/ead/
eLib-dtd
http://sunsite.berkeley.edu

Allgemeines

Charles W. *Bailey :* Scholarly Electronic Publishing Bibliography. Univ. Libraries, Univ. of Houston, Houston, TX 77204-2091. (Wird regelmäßig aktualisiert, dort auch Hinweise auf weitere Projektberichte).
http://info.lib.uh.edu/sepb/sepb.doc (auch html, pdf).

Anne R. *Kenney* and Stephen *Chapman:* Digital Imaging for Libraries and Archives. Dep. of Preservation and Conservation, Cornell Univ. Lib. Ithaca, New York 14853, June 1996.

Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut

Von Peter Exner

Die Motivation, das DFG-Projekt *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Archivalien und Bücher* zu betreiben, liegt in der Bestandserhaltung. Das Archiv- und Bibliotheksgut, das förmlich unter den Fingern der Benutzer zu zerrinnen droht, zu erhalten, ist zugleich Ausgangslage und Problemstellung dieses Projekts. Angesichts des möglichen unwiederbringlichen Verlusts einmaliger Kulturgüter erhält die digitale Technologie für Archive und Bibliothek immense Bedeutung, bietet sie sich doch als Problemlösung an.

Ist sie aber tatsächlich der Lösungsschlüssel für alle Anforderungen der Archivare und Bibliothekare, nicht zuletzt für deren Bestandserhaltungssorgen? Liefert die Digitalisierung vom Zerfall bedrohter Unikate einen veritablen und – vor allem sicheren – Langzeitspeicher, wie ihn diese Institutionen ja zur Erledigung ihrer gesetzlichen Aufgabe benötigen? Stellt die Digitalisierung dabei eine ernstzunehmende Alternative zum bisherigen probaten Speichermedium, dem Mikrofilm, dar oder ist sie eher ein attraktives Zugriffs- und Nutzungsmedium? Ist sie das bedienungsfreundliche Medium, das die Arbeit des Archiv- und Bibliotheksbesuchers erleichtert; stellt sie die Hilfe dar, auf die der Benutzer bei seinen Recherchen per Mausklick gern zurückgreift?

Die angedeuteten Vorfragen lassen sich im Kern auf das Thema zurückführen: Ist das digitale Medium eine Ergänzung oder gar eine Alternative zur Konversionsform Mikrofilm? Zur Klärung dieses Sachverhalts bearbeitete die Landesarchivdirektion Baden-Württemberg das DFG-Projekt *Digitale Konversionsformen*, wie es in Kurzform heißt. In dem vorliegenden Aufsatz werden drei Aspekte dieses Projekts näher vorgestellt: die Projektstruktur und -konzeption, die Arbeits-/Materialgrundlage des Projekts und schließlich die Verfilmung, Direkt- und Filmdigitalisierung als erste, grundlegende Arbeitspakete.

Der Projektstrukturplan mit den Arbeitspaketen

Für eine rationelle Aufbereitung und Abwicklung wurde das Projekt in zehn einzelne Arbeitspakete (AP) zerlegt. Diese wurden wiederum in erforderliche Teilarbeitsschritte untergliedert. Einzelne Arbeitspakete bauen auf einander auf, greifen ineinander oder können teilweise synchron bearbeitet werden.

Dadurch ergaben sich Rationalisierungspotentiale für die tägliche Arbeit und Synergieeffekte in der Aufgabenerledigung.

AP 1: Herstellung geeigneter Mikrofilme
AP 2: Direktdigitalisierung vom Original
AP 3: Digitalisierung vom Mikrofilm
AP 4: Automatische Texterkennung
AP 5: Technische Parameter für den digitalen Zugriff
AP 6: Fachkonzepte zur objekt- und mediengerechten Präsentation
AP 7: Langzeitspeicherung und Migration
AP 8: Ausbelichtung digitaler Konversionsformen auf Mikrofilm (COM)
AP 9: Expertentagung
AP10: Ergebnissicherung

Die Arbeitspakete 1–3 legen den Grundstein, indem sie die Materialgrundlage an analogen und digitalen Konversionsformen bringen, die in den folgenden Arbeitspaketen weiteren Untersuchungen unterzogen werden. Arbeitspaket 4 befaßt sich mit der automatischen Texterkennung (Optical Character Recognition), Arbeitspaket 5 mit den technischen Parametern für den digitalen Zugriff. Im Arbeitspaket 6 werden Fachkonzepte zur objekt- und mediengerechten Präsentation von Archivgut erarbeitet. Diese berücksichtigen erstens Benutzerbedürfnisse, tragen zweitens archivfachlichen Anforderungen Rechnung und nutzen drittens die mit dem digitalen Medium verbundenen Möglichkeiten des Anbietens digitaler Bilder. In Arbeitspaket 7 geht es um Migrationskonzepte zur langfristigen Speicherung und in Arbeitspaket 8 um die Kompatibilität analoger und digitaler Konversionsformen, die den Hintergrund für die Ausbelichtung digitaler Daten auf Mikrofilm (Computer Output on Microfilm) bildet. Die Expertentagung (Arbeitspaket 9) dient dem wissenschaftlichen Gespräch, und die Ergebnissicherung (Arbeitspaket 10) erfolgt in Form geeigneter Projektdokumentationen (Berichte, Internetangebot) und Publikationen.

Die *Warenkörbe* als Arbeitsgrundlage

Der *große Warenkorb*

Arbeits- und Materialgrundlage für alle Arbeitspakete, insbesondere für die Verfilmung und Digitalisierung (Direktdigitalisierung und Filmdigitalisierung), waren typische Vertreter einzelner Gattungen von Archiv- und Bibliotheksgut, die in einem sogenannten großen Warenkorb zusammengestellt worden waren. Die Auswahl erfolgte im Einvernehmen mit den baden-württembergischen Staatsarchiven, dem Hauptstaatsarchiv Stuttgart (HStAS) und dem Generallandesarchiv Karlsruhe (GLAK), sowie in Zusammenarbeit mit der Württembergischen Landesbibliothek (WLB).

Als vorbereitende Maßnahme wurde für den Warenkorb ein Numerierungssystem eingeführt, mit dem jedes Objekt identifizierbar ist und das Bestandteil der im Rahmen der Verfilmung und Digitalisierung vergebenen Kennungen ist. So steht *LK 309 205* für:

LK: Landesarchivdirektion Baden-Württemberg Konversionsformen (*großer Warenkorb*)
309 (die ersten drei Ziffern): *Warenkorbnummer* des einzelnen Objekts (z.B. Akte, Buch)
205 (die letzten drei Ziffern): laufende Aufnahme.

Die Auswahl von geeigneten Archivalien aus den Beständen baden-württembergischer Staatsarchive erfolgte nach folgenden Kriterien:

- Die Vertreter der einzelnen Archivaliengattungen sollten inhaltlich aussagekräftig und optisch ansprechend sein.
- Ebenfalls wurde darauf geachtet, daß die Repräsentanten der jeweiligen Gattung typisch sind; archivwissenschaftlich sollten sie sowohl das Allgemeine wie das Besondere ihrer Gattung widerspiegeln.[1]
- Wo immer möglich, wurden anstehende Jubiläen aus dem Sprengel der baden-württembergischen Staatsarchive berücksichtigt, z.B. 150 Jahre Revolution 1848/49 im deutschen Südwesten, 850 Jahre Kloster Maulbronn, 600 Jahre Verbindung der Häuser Württemberg und Mömpelgard (Montbéliard) oder 80 Jahre Revolution 1918/19.

Quer zu den Gattungen lassen sich die rund 150 Stücke auch nach folgenden Aspekten charakterisieren:

- nach dem Beschreibstoff: vertreten sind Pergament, Papier unterschiedlichster Art und Güte (Photopapier oder säure- und holzschliffhaltiges Massenpapier) sowie Glasplatten und Wachssiegel.
- nach der äußeren Form: es wurden vorrangig zweidimensionale, aber auch dreidimensionale Vorlagen (Siegel und Typare) ausgesucht.
- nach der Farbigkeit: vertreten sind sowohl zweifarbige (schwarz-weiß) als auch mehrfarbige Vorlagen.

Die Einzelstücke mit Signatur sind im Internet-Projektangebot online abrufbar,[2] deshalb wird hier nur eine systematische Aufzählung der Gattungen geboten.

---

[1] Siehe dazu Johannes *Papritz*: Archivwissenschaft, Bd. 2, Teil II,2: Organisationsformen des Schriftguts in Kanzlei und Registratur. Marburg 1976. S. 284–288.

[2] URL: *http://www.lad-bw.de/digpro/texte/warenkor.htm*.

Aus 18 Archivaliengattungen wurden folgende typische Vertreter ausgesucht:

1. Akten (darunter frühneuzeitliche Akten des 16.–18. Jh. und Akten des 19. Jh. sowie maschinenbeschriebene moderne Akten des 20. Jh.)
2. Amtsbücher
3. Drucke
4. Flugschriften bzw. Maueranschläge
5. Karten
6. Kupferstiche
7. Lithographien
8. Notariatsinstrumente (darunter klassische Notariatsinstrumente des 15.–17. Jh. sowie ein Libell des 18. Jh.)
9. Photographien
10. Plakate (aus der NS- und Nachkriegszeit, zur Entstehung des Südweststaates und der bundesrepublikanischen Parteienlandschaft)
11. Pergament-Rotulus
12. Siegel und Typare
13. Telegramme
14. Urkunden (nach der klassischen Diplomatik unterteilt in Kaiser-, Königs-, Papst- und Privaturkunden)
15. Wappen und Wappenbücher
16. Zeitungen
17. Testtafeln

Aus den Beständen der Württembergischen Landesbibliothek wurden folgende typische Vertreter einzelner Gattungen von Bibliotheksgut ermittelt:

1. Buchdrucke (15.–20. Jh.)
2. Glasplatten
3. Zeitungsausschnitte
4. Kupferstiche
5. Tonlithographien
6. Photoabzüge
7. Radierungen
8. Federzeichnungen
9. Stiche (Mischtechnik)
10. Radierungen (aquarelliert)
11. Bleistiftzeichnungen
12. Aquarelle
13. Schabkunstblätter

Der *kleine Warenkorb*

Bei der Arbeit mit dem *großen Warenkorb* hat sich gezeigt, daß für bestimmte vergleichende Untersuchungen ein weiterer sog. kleiner Warenkorb notwendig ist. Der *kleine Warenkorb* ist keine Untermenge des *großen*, sondern eine eigenständige Materialauswahl bestehend aus kassierten Akten, Zeitungsseiten, Buchseiten und ausgewählten Testmuster-Seiten unterschiedlicher Provenienz. Es handelt sich dabei bewußt um kein zur dauerhaften Aufbewahrung bestimmtes Archivgut mit unwiederbringlichem Wert. Insgesamt umfaßt der *kleine Warenkorb* 35 Objekte (31 Schriftstücke/Abbildungen und vier Testmusterseiten). Eine nähere Beschreibung der einzelnen Objekte ist im Internetangebot abrufbar.[3]

Beim *kleinen Warenkorb* wurde ebenfalls eine Verfilmung, Direktdigitalisierung und Filmdigitalisierung vorgenommen. Seine Objekte wurden insbesondere für die Erprobung der automatischen Texterkennung (OCR)[4] und des COM-Verfahrens[5] verwendet aufgrund der besonderen Eignung der Objekte für eine bitonale Verarbeitung. Schließlich ist diese Zusammenstellung von Testmaterial auch für Einzeluntersuchungen in Verbindung mit einem Flachbettscanner geeignet.

Für den *kleinen Warenkorb* wurde analog zum großen ein medienübergreifendes Numerierungssystem eingeführt, das bei der Verfilmung und Digitalisierung Bestandteil der Kennungen ist. So steht *LKW13A* bzw. *LKW13B* für:

LKW: Landesarchivdirektion *kleiner Warenkorb*
13A: Objekt 13 Vorderseite
13B: Objekt 13 Rückseite

Verfilmung und Digitalisierung

Herstellung geeigneter Mikrofilme

Bei der Ausschreibung für die Anfertigung geeigneter analoger Konversionsformen wurde ein detailliertes Anforderungsprofil entworfen, das die Grundlage der Auftragsvergabe an einen Dienstleister bildete. Zugrundegelegt wur-

---

3 URL: *http://www.lad-bw.de/digpro/texte/warenkor.htm*.

4 OCR= Optical Character Recognition. Siehe dazu in diesem Band den Beitrag von Thomas *Fricke* und Gerald *Maier*: Automatische Texterkennung bei digitalisiertem Archiv- und Bibliotheksgut. S. 201.

5 *Computer Output on Microfilm*-Verfahren. Siehe dazu in diesem Band den Beitrag von Hartmut *Weber:* Langzeitspeicherung und Langzeitverfügbarkeit digitaler Konversionsformen. S. 325.

den dabei die Erkenntnisse der Arbeitsgruppe *Digitalisierung* des Unterausschusses *Bestandserhaltung* des Bibliotheksausschusses der Deutschen Forschungsgemeinschaft (DFG).[6] Danach ist der Mikrofilm hinsichtlich seiner Wirtschaftlichkeit, Haltbarkeit und Wiedergabequalität das geeignetste Speichermedium und damit die effizienteste Alternative zur herkömmlichen Konservierung.

Die Mikroverfilmung der Warenkorbstücke soll hier an dieser Stelle als Arbeitspaket gemäß der Projektkonzeption kompakt und knapp skizziert werden. Die Übersicht über die einzelnen Teilarbeitsschritte dieses ersten Arbeitspakets des Projekts erfolgt in Gestalt einer stichwortartigen Auflistung.

AP1: Herstellung geeigneter Mikrofilme

- TA 1.1: Festsetzung der Gattungen von Archivalien und Büchern
- TA 1.2: Ermittlung von Vertretern der einzelnen Gattungen von Archivalien und Büchern
- TA 1.3: Erstellen eines Warenkorbs im Benehmen mit HStAS, GLAK und WLB
- TA 1.4: Analyse der Anwendbarkeit des Anforderungsrasters der AG „Digitalisierung“
- TA 1.5: Erarbeitung alternativer Anforderungen und Lösungen für Stücke, für die TA 1.4 nicht trägt
- TA 1.6: Erarbeitung von Vorgaben für Aufnahmedokumentation und Dateiorganisation
- TA 1.7: Festsetzung des Verfilmungsprogramms
- TA 1.8: Ausarbeitung einer Ausschreibung
- TA 1.9: Markterkundung
- TA 1.10: Auftragsvergabe an Dienstleister; intensive Begleitung bei Durchführung des Auftrags
- TA 1.11: Kontrolle auf a) Vollständigkeit b) Qualität nach DIN
- TA 1.12: Gegenüberstellung von Kosten und Aufwand
- TA 1.13: Ergebnisdokumentation

Abb. 1: Strukturplan Projekt *Digitale Konversionsformen,* Arbeitspaket 1.

Die rund 150 Stücke des *großen Warenkorbs* wurden bei einem Dienstleister komplett verfilmt, ebenso der gesamte *kleine Warenkorb*. Die Verfilmung orientiert sich bei der Filmauswahl und -qualität, Aufnahmetechnik sowie Orga-

[6] Siehe dazu Marianne *Dörr* und Hartmut *Weber:* Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft. In: Zeitschrift für Bibliothekswesen und Bibliographie 44 (1997) Heft 1 S. 58–64 und online unter: *http://www.lad-bw.de/lad/dfgdigh1.htm.* Zur objektgerechten Mikroverfilmung und den verschiedenen Mikroformen siehe auch Hartmut *Weber:* Verfilmen oder Instandsetzen? Schutz- und Ersatzverfilmung im Dienste der Bestandserhaltung. In: Bestandserhaltung in Archiven und Bibliotheken. Hg. von Hartmut *Weber* (Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg A 2). Stuttgart 1992. S. 103–111.

nisation und Aufnahmedokumentation an den Richtlinien der Arbeitsgruppe *Digitalisierung*. Da die Verfilmung auch im Hinblick auf eine spätere Digitalisierung vom Film erfolgte, wurde dafür gesorgt, daß jede Filmaufnahme für die Digitalisierung nutzbar ist. Gegenüber einer reinen Sicherungsverfilmung wurde bei den Rollfilmen auf eine gleichmäßige und exakte Positionierung der Objekte für eine eventuelle automatische Seitentrennung geachtet. Die Filmaufnahmen wurden mit einer Kennung in Form einer Titeleinbelichtung versehen, die auf der Numerierung der Warenkörbe basiert. Alle Filme wurden einer Qualitätsprüfung nach DIN unterzogen.

Verfilmung auf 35mm Rollfilm

Die Mehrzahl der Objekte wurde auf unperforierten 35mm Rollfilmen mit Blip-Einbelichtung nach DIN/ISO 1699 bzw. DIN 19057 in den Bildlagen 1A, 1B und 2B verfilmt. Die Verarbeitung der entstandenen Filme erfolgte unter Berücksichtigung der einschlägigen Normen (v.a. DIN 6199). Von den 35mm-Aufnahmen wurden pro Vorlage eine Originalaufnahme und drei Duplikate für die Weiterverarbeitung erstellt.

Aus dem *großen Warenkorb* wurden kontrastreiche Vorlagen mit relativ homogener Papier- und Schriftdichte wie Akten, Amtsbücher, Drucke und Telegramme mit einem konventionellen AHU-Mikrofilm für negative Schwarzweiß- Strichaufnahmen (Agfa Copex HDP13 35mm) aufgenommen.[7] Vorlagen mit Grautönen, z.B. Photographien, Lithographien und Plakate, wurden mit einem Halbtonmikrofilm für positive schwarzweiß-Halbtonaufnahmen (Kodak Direct Duplicating 2468 35mm) erfaßt. Zwei- bzw. dreidimensionale Farbvorlagen wurden mit einem Farbfilm für positive Color-Aufnahmen (Ilfochrome Micrographic CMP 35mm) aufgenommen. Dies betraf alle Plakate und Wappen bzw. Siegel und Typare. Am Anfang jeder Filmrolle wurde eine standardisierte Farb-Testtafel (Colorchecker) mit aufgenommen. Die Filme wurden durchgezählt und je nach Institution (Archiv/Bibliothek) getrennt. Stücke mit einem umgebenden Tableau, wie sie Kupferstiche oder Photographien in Form eines Passepartout aufweisen, wurden stets im Vollformat aufgenommen, um bei der anstehenden Filmdigitalisierung nicht unnötige Speichermenge zu verbrauchen.

---

[7] AHU (Anti-Halation-Undercoated)-Mikrofilme sind mit einer sogenannten Lichthofschutzschicht versehen, wodurch die Schärfe der photographischen Abbildung verbessert wird. AHU-Filme sind panchromatisch, kontraststeigernd und eignen sich vorzugsweise für Strichvorlagen. Vgl. dazu Barbara *Keimer:* Mikroverfilmung von Büchern als bestandserhaltende Maßnahme (dbi Materialien 154). Berlin 1997. S. 15.

Filmrolle	Objekt-Nr. der Vorlage aus dem *Großen Warenkorb*	Filmtyp
1	310-305	Agfa Copex HDP 13
2	305 (Forts.), 307, 309 310, 312	Agfa Copex HDP 13
3	302, 306	Ilfochrome Micrographic CMP
4	311, 313, 321, 126, 127	Ilfochrome Micrographic CMP
5	314, 316, 317, 126, 127	Agfa Copex HDP 13
6	301, 303–306	Kodak Direct Duplicating 2468
7	306 (Forts.)–308	Kodak Direct Duplicating 2468
8	309–312	Kodak Direct Duplicating 2468
9	313, 315–317, 319–321, 126, 127	Kodak Direct Duplicating 2468
10	1–127	Agfa Copex HDP 13
11	1–127	Agfa Copex HDP 13
12	1–127	Ilfochrome Micrographic CMP
13	1–127	Kodak Direct Duplicating 2468
14	1–127	Kodak Direct Duplicating 2468
15	1–127	Kodak Direct Duplicating 2468

Abb. 2: Übersicht über die Verfilmung des *großen Warenkorbs.*

Der *kleine Warenkorb* wurde komplett mit Testzeichen sowohl auf 35mm AHU-Negativ (Agfa Copex HDP13 35mm) als auch Halbton-Positiv-Mikrofilm (Kodak Direct Duplicating 2468 35mm) verfilmt.

Verfilmung auf Makrofiches

Karten, Graphiken und Kupferstiche wurden aufgrund ihres Formats als Farb- bzw. Halbton-Makrofiches (105mm, Format DIN A6) verfilmt. Die Vorlagen wurden formatfüllend aufgenommen, wenn erforderlich wurden sie um 90° gedreht; der Kopf der Aufnahme befand sich dann links.

Graustufenvorlagen wurden mit einem Film für positive Schwarzweiß-Halbtonaufnahmen (Kodak Direct Duplicating 2468 105mm), Farbvorlagen mit einem Film für positive Color-Aufnahmen (Ilfochrome Micrographic CMP 105mm) zusammen mit einem Farbkeil aufgenommen.

Die Titelleiste mit Titelaufnahme wurde dabei dreigeteilt; links: die laufende Nummer, korrespondierend mit der Warenkorbnummer, darunter die jeweilige Institution; mittig: die Titelkurzform; rechts: das Aufnahmedatum und die fortlaufende Aufnahmezahl. Von den Makrofiches wurden pro Vorlage eine Erstaufnahme und fünf weitere Aufnahmen erstellt.

Qualitätskontrolle der Verfilmung

Die 15 Rollen Mikrofilm (5 Rollen AHU-, 7 Rollen Halbton- und 3 Rollen Farbfilm) und 122 Makrofiches des *großen Warenkorbs* sowie die 2 Rollen (je 1 Rolle AHU- und Halbtonfilm) des *kleinen Warenkorbs* wurden einer Vollständigkeits- und Qualitätskontrolle nach DIN unterzogen. Dabei lagen alle Filme in puncto Auflösung über der Norm; mit einer Dichte von 1,0 (+/- 10%) entsprachen sie der Norm voll und ganz. Die Auflösungskennzahl erreichte 14 bis 16 des ISO Testzeichens Nr. 2;[8] das entspricht 160 bis 180 Linienpaaren pro Millimeter auf das Original bezogen. Gemessen am Qualitätsindex (QI)[9] lagen diese Werte besser als 8, also im Bereich der höheren Qualität. Die Halbtonwiedergabe war bei AHU-Filmen vergleichsweise besser als bei digitalen Bildern mit 16 Graustufen; bei speziellen Halbtonfilmen war sie weit ansehnlicher als bei einer Digitalisierung mit 256 Graustufen.

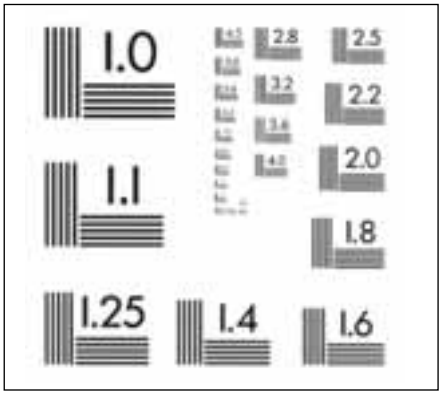


Abb. 3: ISO-Testzeichen Nr. 2 (Linienpaare/mm).

---

8 ISO-Testzeichen Nr. 2 besteht aus zwei zueinander senkrecht stehenden Liniengruppen aus je fünf dunklen und fünf hellen Linien (Linienpaare). Die Kennzahl gibt die dargestellten Linienpaare je mm an. Siehe dazu DIN-Taschenbuch 154. Publikation und Dokumentation 2. Normen. Hg. vom DIN-Institut für Normung e.V., Berlin u.a. [4]1996. S. 76–84.

9 Zum Qualitätsindex (QI) für die Beurteilung von Mikrofilmen siehe beispielsweise RLG Preservation Microfilming Handbook. Ed. by Nancy E. *Elkington*. Mountain View. California 1992. P. 171–174.

Die Verfilmung dreidimensionaler Vorlagen (Siegel und Typare) erwies sich bei objektgerechter Ausleuchtung als unproblematisch. Durch seitlich einfallendes Licht konnte ein guter Kontrast und eine lupenreine Tiefenschärfe erzielt werden, womit die Plastizität dreidimensionaler Vorlagen betont wurde. Dies gilt für Blei- und Wachssiegel sowie Typare, aber auch für aufgedrückte Papiersiegel, deren Wappen klar erkennbar und deren Umschrift kenntlich lesbar sind. Bei Büchern wurde immer auch der Buchrücken bzw. bei Akten der Aktendeckel sowie alle Leerseiten mit erfaßt, um die Vorlage in möglichst authentisch wiederzugeben. Bei Urkunden und Siegeln wurde immer auch die Rückseite mit aufgenommen, um Gegensiegel sowie Adressen und Empfängervermerke mit zu erfassen. Großformatige, vor allem längliche Karten wurden im Querformat abgebildet.

Dagegen war die Wiedergabequalität von Bleistiftzeichnungen auffallend gering. Hier war vor allem ein mangelhafter Kontrast zu konstatieren. Auf AHU-Filmen waren im Gegensatz zu Halbtonfilmen Photographien auf Zeitungspapier nicht mehr einwandfrei erkennbar.

Direktdigitalisierung vom Original

Auch bei der Digitalisierung vom Original prüfte das Projekt die Anwendbarkeit der Empfehlungen der Arbeitsgruppe *Digitalisierung* des Unterausschusses *Bestandserhaltung* des Bibliotheksausschusses der Deutschen Forschungsgemeinschaft (DFG).[10]

Grundlage für die Direktdigitalisierung waren die verschiedenen *Warenkörbe*, die sich aus Vertretern verschiedener Gattungen von Archiv- und Bibliotheksgut zusammensetzten.

Vorgenommen wurde die Direktdigitalisierung der einzelnen Warenkorbobjekte durch einen Dienstleister. Im Unterschied zur Mikroverfilmung wurde bisher nicht alles komplett digitalisiert. Vielmehr wurden in Auswahl – je nach technischer Anforderung (Farbe, Grautöne, Auflösung) – einzelne Seiten der Vorlagen erfaßt. Während bei der Verfilmung doppelseitig aufgenommen wurde, wurde bei der Direkt- wie auch bei der Filmdigitalisierung einseitig aufgenommen.

Die Direktdigitalisierung erfolgte formatfüllend bis an die Grenze der Chipgröße von knapp 4500 x 3500 Pixels. Digitalisiert wurde immer mit der maximalen Auflösung und der bestmöglichen Größeneinstellung, um die Chipfläche optimal auszunutzen. Aufgenommen wurde mit Hilfe einer CCD-Kamera mit Spezialobjektiven, die aus der Mikroverfilmung stammen.

---

10 Siehe dazu *Dörr/Weber*, wie Anm. 6, S. 73–75.

Bei der maximalen Auflösung von 4491 x 3480 Bildpunkten ergaben sich – bezogen auf das Original in mm – folgende Auflösungen (in dpi):

Deutsche Industrie Norm (DIN)	Original	(in mm)	Auflösung in dpi
DIN A0	1189	841	96
DIN A1	841	594	136
DIN A2	594	420	192
DIN A3	420	297	272
DIN A4	297	210	384
DIN A5	210	148	543
DIN A6	148	105	771

Von den rund 150 Archivalien und Büchern des *großen Warenkorbs* wurden 186 Graustufenscans und 190 Farbscans erzeugt. Bei Farbaufnahmen wurde die Kamera auf den *Macbeth Colorchecker* kalibriert, der es ermöglicht, Farbfehler eines farbigen Sicherungsfilms weitgehend zu neutralisieren, wenn bei dessen Aufnahme ein solcher Colorchecker mitverfilmt wurde. Die Schärfeneinstellung wurde von einer Autofocusfunktion elektronisch überprüft und justiert.

Die bei der Direktdigitalisierung erzeugten Daten wurden verlustfrei und unkomprimiert im TIF-Format abgelegt und auf jeweils zwei CD-Rs ausgegeben. Durch die Bildbearbeitung, bei der überflüssige Aufnahmen weggeschnitten wurden, konnte die Datenmenge reduziert werden.

Auflösungsstufen in Bildpunkten			Größe der Dateien in Megabyte (unkomprimiert)		
Größe	X	Y	Farbe	Graustufen	Bitonal
A	4491	3480	46,89	15,63	1,95
B	3992	3900	34,73	11,58	1,45
C	2994	2320	20,84	6,95	0,87
D	1996	1450	8,68	2,89	0,36
E	1497	1160	5,21	1,74	0,22
F	998	774	2,32	0,77	0,10
G	749	580	1,30	0,43	0,05
H	499	387	0,58	0,19	0,02

Die Bildbearbeitung der direkt generierten Scans erfolgte mit dem Bildbearbeitungsprogramm Adobe Photoshop 4.0, das u.a. eine Stapelbearbeitung zuläßt, die ein Bearbeiten aller Bilder mit gleicher Konfiguration durch ein vorab definiertes Makro ermöglicht. Die digitalen Bilder wurden einer Farb-,

Kontrast-, Helligkeits- und Schärfenkorrektur unterzogen. Die Bildbearbeitung folgte ausschließlich dem Ziel, eine möglichst getreue Wiedergabe des Originals zu erzeugen.

Die unterschiedlichen Aufnahmeverfahren, also Verfilmung, Direktdigitalisierung und Filmdigitalisierung, sind notwendig, um in Vergleichsuntersuchungen die Kompatibilität analoger und digitaler Konversionsformen bzw. analoger und digitaler Konversionsformen erster und zweiter Generation zu ermitteln. Weiteren Aufschluß werden hierbei Versuche mit der Ausbelichtung digitaler Daten auf Mikrofilm (COM-Verfahren) erbringen.

Die digitalen Bilddateien werden in einem weiteren Schritt einer Qualitätsprüfung unterworfen und in einer Bilddatenbank mit Thumbnail-Generierung katalogisiert. Für die Qualitätskontrolle werden in Anlehnung an die Standards bei der Mikroverfilmung Kriterien erarbeitet.

Digitalisierung vom Mikrofilm

Die Filmdigitalisierung diente dem Ziel, einerseits die Kompatibilität und Qualität von analogen und digitalen Konversionsformen zu vergleichen, andererseits wurden aber auch Qualität und Wirtschaftlichkeit der Filmdigitalisierung mit dem Verfahren der Direktdigitalisierung verglichen. Auch hier wird die detaillierte Gliederung dieses Arbeitspakets mit seinen Teilarbeitsschritten geboten:

AP 3: Digitalisierung vom Mikrofilm

- TA 3.1: Ermittlung der Anwendbarkeit der Empfehlungen der AG „Digitalisierung“
- TA 3.2: Erarbeitung von Qualitätsanforderungen für Digitalisierung von Mikrofilmen, für die TA 1 nicht trägt (ggf. im Benehmen mit erfahrenen Dienstleistern)
- TA 3.3: Festsetzen der Erschließungsmerkmale, die beim Digitalisieren mit erfaßt werden sollen (im Benehmen mit erfahrenen Dienstleistern)
- TA 3.4: Festsetzung des Digitalisierungsauftrags
- TA 3.5: Markterkundung
- TA 3.6: Auftragsvergabe an mehrere Dienstleister mit unterschiedlicher Geräteausstattung; intensive Begleitung des Auftrags
- TA 3.7: Erarbeitung eines Konzeptes für visuelle Qualitätskontrolle digitaler Konversionsformen und Festsetzung eines Beurteilungsrasters
- TA 3.8: Kontrolle auf a) Vollständigkeit, b) Qualität
- TA 3.9: Kostenvergleiche zwischen den verschiedenen Dienstleistern
- TA 3.10: Kosten und Nutzen der digitalen Erschließungsinformation
- TA 3.11: Ergebnisdokumentation

Abb. 4: Strukturplan Projekt *Digitale Konversionsformen,* Arbeitspaket 3.

Bei der Digitalisierung von Mikrofilmen prüfte das Projekt ebenfalls die Anwendbarkeit der Empfehlungen der Arbeitsgruppe *Digitalisierung* des Unterausschusses *Bestandserhaltung* des Bibliotheksausschusses der Deutschen Forschungsgemeinschaft (DFG).[11]

Grundlage für die Filmdigitalisierung bildete die Mikroverfilmung der *Warenkörbe*. Zur Identifikation einzelner Objekte wurde basierend auf der Warenkorbnumerierung ein System von Kennungen geschaffen, das die Verwendung der Objekte in Bilddatenbanken ermöglicht.

Marktanalyse und Ausschreibung

Zur Erledigung des Arbeitspakets wurde entsprechend der Auftragsvergabe bei der Verfilmung ein Digitalisierungsauftrag festgesetzt, der ein detailliertes Anforderungsprofil für die Digitalisierung von AHU-, Halbton- und Farb-Mikrofilmen sowie von Makrofiches in verschiedenen Aufnahmequalitäten enthält. Gefordert wurden folgende Leistungen:

- Digitalisierung von AHU-Mikrofilmen 35mm unperforiert negativ:
  - bitonal (1 Bit): a) 400 dpi und b) 600 dpi
  - 16 Graustufen (4 Bit): a) 400 dpi und b) 600 dpi
- Digitalisierung von Halbton-Mikrofilmen 35mm unperforiert positiv:
  - bitonal (1 Bit): a) 400 dpi und b) 600 dpi
  - 16 Graustufen (4 Bit): a) 300 dpi und b) 400 dpi
  - 256 Graustufen (8 Bit): a) 250 dpi und b) 350 dpi
- Digitalisierung von Farb-Mikrofilmen 35mm unperforiert positiv:
  - bestmögliche Qualität bis 400 dpi
- Digitalisierung von Makrofiches (105x148mm), Mikroplanfilme DIN A 6 mit einer Aufnahme im Format 90x127mm:
  - bitonal (1 Bit) mit der bestmöglichen Aufnahmequalität bis 600 dpi
  - 16 Graustufen (4 Bit) mit der bestmöglichen Aufnahmequalität bis 400 dpi
  - 256 Graustufen (8 Bit) mit der bestmöglichen Aufnahmequalität bis 350 dpi
  - Farbe (24 Bit) mit der bestmöglichen Aufnahmequalität bis 400 dpi

Die digitalen Bilder sollten im TIF-Format gespeichert und auf CD-R-Datenträger geliefert werden.

Diese Ausschreibung wurde allen dafür in Frage kommenden Anbietern in Deutschland und dem europäischen Ausland zugeschickt, eine Anfrage ging

---

[11] Ebd., S. 64–70.

nach Übersee (Australien). Insgesamt wurden damit zehn Serviceunternehmen angesprochen; eingegangen sind sechs Angebote.[12]

Dabei führten alle Dienstleister an, für die Nennung konkreter Verarbeitungspreise Musterfilme zu benötigen, deren Qualität zu testen sei. Alle Dienstleister, außer einer Firma, baten um ein vertiefendes Gespräch, z.B. über Mengengerüst, sowie über Testläufe mit Testfilmen und über die eingesetzte Hard- und Software, um die Angebote weiter konkretisieren zu können.

Filmdigitalisierung des großen Warenkorbs

Die Filmdigitalisierung des *großen Warenkorbs* erfolgte in Graustufen (8 Bit) bzw. in Farbe (24 Bit) durch einen Dienstleister mit Hilfe eines speziellen Aufsicht-Filmscanners mit CCD-Chip der Firma Kontron und einer Durchlichteinrichtung.

Der CCD-Chip wird bei der Aufnahme mit Piezo-Elementen bewegt (sogenanntes *Schüttelverfahren*), so daß neben den bereits erfaßten weitere Bildpunkte aufgenommen werden können. Der Chip bewegt sich dadurch in drei Richtungen in zweidimensionaler Ebene: in die Länge, Breite und Diagonale.

Bei der maximalen Auflösung von 4491x3480 Bildpunkten ergaben sich – bezogen auf das Original in mm – folgende Auflösungen (DPI):

Deutsche Industrie Norm (DIN)	Original (mm)		DPI
Makrofiches (DIN A6)	148	105	771
35mm Rollfilm Bildlage 1A/2B	42	32	2716
35mm Rollfilm Bildlage 1B	32	21	3565

Filmdigitalisierung des kleinen Warenkorbs

Die 35 Objekte des *kleinen Warenkorbs* wurden von verschiedenen Dienstleistern in unterschiedlichen Auflösungen (300–600 dpi, bezogen auf die Vorlage) und Farbtiefen (1–8 Bit) gescannt. Die digitalen Master wurden in den Dateiformaten TIF und TIF-G4 abgespeichert.

Die Filmscans des *kleinen Warenkorbs* dienten als Grundlage für Fragen der Bildqualität, Untersuchungen mit verschiedenen Dateiformaten und Dateikomprimierungsverfahren für digitale Präsentationsmaster und für die Erprobung von OCR-Verfahren.

---

12 Zur unterschiedlichen Preisgestaltung und den verschiedenen Preisspannen für Digitalisierungsleistungen siehe in diesem Band den Beitrag von Gerald *Maier* und Peter *Exner:* Wirtschaftlichkeitsüberlegungen bei der Digitalisierung von Archiv- und Bibliotheksgut. S. 223.

Am *kleinen Warenkorb* wurden unterschiedliche Filmscan-Technologien erprobt, zum einen die schon bei der Digitalisierung des *großen Warenkorbs* eingesetzte Aufsichtscantechnik mit Durchlichteinrichtung, zum anderen spezielle geschlossene Mikrofilmscanner mit weitgehender Aufnahmeautomatisierung.

Tests mit folgenden speziellen Mircofilmscannern wurden dabei von verschiedenen Dienstleistern durchgeführt:

- Sunrise ProScan III mit CCD-Kameratechnik[13]: Auflösung bis 300 dpi, 1 Bit und 8 Bit,
- Sunrise ProScan IV mit CCD-Kameratechnik: Auflösung bis 600 dpi, 1 Bit und 8 Bit,
- Fujix MS 6000: Auflösung bis 400 dpi, 1 Bit und 8 Bit,
- Bell + Howell ABR 3000D: Auflösung bis 400 dpi, 1 Bit.

[13] URL: *http://www.sunriseimg.com.*

Qualität, Bearbeitung und Präsentation digitaler Bilder

Von Gerald Maier

Ein Schwerpunkt des Projekts *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Bücher und Archivalien* lag bei der Qualitätsbeurteilung, Bearbeitung und Präsentation digitaler Formen von Archiv- und Bibliotheksgut. Im Rahmen eines Digitalisierungs-Workflow sind neben dem eigentlichen Scanvorgang Bildverwaltung, Qualitätskontrolle, *Image Enhancement* und Aufbereitung digitaler Bilder für die Präsentation wesentliche Schritte.

Sechs Bereiche, die eng miteinander in Verbindung stehen, sollen hier vorgestellt werden:

1. der Einsatz spezieller Standard-Software für Bildbearbeitung und digitalen Zugriff,
2. die Verwaltung digitaler Konversionsformen,
3. die Qualität digitaler Konversionsformen,
4. die geeigneten Dateiformate für digitale Konversionsformen (Abbildungsqualität, Komprimierung, Dateigröße),
5. *Image Enhancement* und Bildaufbereitung und
6. die objekt- und mediengerechte Präsentation von Archiv- und Bibliotheksgut.

Als Testmaterial dienten ausgewähltes Archiv- und Bibliotheksgut, zusammengestellt in sogenannten *Warenkörben.*[1] Dabei wurden Testreihen mit verschiedenen Dateiformaten, Farbtiefen und unterschiedlichen Kompressionsraten durchgeführt. In diesem Zusammenhang wurde auch die Qualität direkt- und filmdigitalisierter Formen miteinander verglichen und die Möglichkeiten einer Qualitätssteigerung durch Bildbearbeitung (*Image Enhancement*) getestet. Die Ergebnisse können online an drei Testmodellen mit aufbereiteten Objekten des *großen Warenkorbs* (Wachssiegel, farbige Wappendarstellung, Zeitung) betrachtet werden.[2] Außerdem wurden verschiedene Verfahren er-

---

[1] Zu den *Warenkörben* siehe in diesem Band den Beitrag von Peter *Exner:* Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut. S. 113.

[2] URL*: http://www.lad-bw.de/digpro/graf.htm.*

probt, mit denen das Bildmaterial für Präsentationen aufbereitet werden kann. Dazu gehörte auch der Test unterschiedlicher Software wie Bildbearbeitungsprogramme, Viewer- und Browser-Software und Entwicklungswerkzeuge. Von besonderer Bedeutung waren dabei automatisierte Stapel-Verfahren (Batch-Verfahren), die eine rationelle Mengenverarbeitung erlauben.

1. Einsatz spezieller Standard-Software für Bildbearbeitung und digitalen Zugriff

Qualitätskontrolle, Bildbearbeitung, Bildverwaltung und Aufbereitung von digitalen Bildern für Präsentationen erfordern den Einsatz spezieller Software. Daher wurden verschiedene Softwareprogramme beschafft und auf 8x86 PCs unter dem Betriebssystem Microsoft Windows NT 4.0 getestet.

Bildbearbeitung/Viewer/Web-Browser

Photoshop 4.0/5.02 (Adobe)

Für die Bildbearbeitung und das Image Enhancement wurde in erster Linie das professionelle Bildverarbeitungsprogramm Adobe Photoshop eingesetzt. Version 5.02 bietet vor allem eine verbesserte Funktionalität im Batch-Betrieb und bei der Makroerstellung.

Das Programm wurde auch zur Konvertierung von TIFF-Dateien in JPEG-Dateien für Internetpräsentationsmodelle verwendet mit gleichzeitiger Dateigrößenreduktion, Tonwertkorrektur und Schärfekorrektur. Über Makros und Batch-Betrieb wurden auch automatische Seitentrennungen von doppelseitigen Images vorgenommen. Mit Photoshop wurden für Standardabläufe der Bildaufbereitung verschiedene Makros erstellt.

PhotoImpact 3.0/4.2 (Ulead Systems)

Hierbei handelt es sich um ein Bildverarbeitungsprogramm mit besonderen Funktionen für die Aufbereitung von Bildmaterial für Online-Präsentationen. Das Programm bietet eine umfangreiche Unterstützung der internetrelevanten Bildformate JPEG, PNG und GIF, automatisierende Bearbeitungsfunktionen (Stapelverarbeitung) und ein Zusatzmodul *SmartSaver* für eine visuelle Kontrolle der Bildqualität unter Berücksichtigung der Dateigröße bei den Formaten GIF, JPEG und PNG. Leider ist das Modul *SmartSaver* nicht für den Stapelbetrieb geeignet aufgrund unzulänglicher Makro- und Batchfunktionalität.

ThumbsPlus 3.3 und 4.0x (Cerious Software)

Das Programm ThumbsPlus wurde außer zur Qualitätskontrolle und Bildkatalogisierung auch zur Batch-Verarbeitung von Images eingesetzt. Das Programm eignet sich besonders für die Aufbereitung von digitalen Mastern zu Präsentationsformen. Gut ist die Unterstützung des leistungsfähigen PNG-Dateiformats.

Das Update von 3.3 auf die Version 4.0 brachte vor allem Verbesserungen im Batch-Betrieb und eine allgemeine Performance-Steigerung.

JPEG Optimizer 3.0x (XATech)

Das Programm ermöglicht die Optimierung von JPEG-Dateien mit visueller Kontrolle der Veränderungen bezüglich Komprimierungsfaktor und Dateigröße. Zusätzlich werden Möglichkeiten geboten, interaktiv bestimmte Bildbereiche mehr oder weniger zu komprimieren. So können z.B. Bereiche mit Text in einem Image für eine bessere Lesbarkeit weniger stark komprimiert werden, der restliche Teil dagegen stärker.

Negativ für eine Mengenbearbeitung ist die schlechte Performance des Programms.

DeBabelizer Pro 4.51 (Equilibrium)

DeBabelizer Pro ist ein mächtiges Konvertierungswerkzeug für verschiedene Bilddateiformate mit umfangreichen Möglichkeiten der Farbreduktion (Einsatz verschiedener Farbpaletten und Ditheringmethoden). Die Stärke des Programms liegt in der Massenbildbearbeitung und in den vielfältigen Bildbearbeitungsmöglichkeiten, die über eine Script- und Makrosprache während des Konvertiervorgangs angewandt werden können.

Für das Programm DeBabelizer Pro wurden vom Projektteam für die Bildaufbereitung verschiedene Makros erstellt. Nachteilig für eine lückenlose Batchbearbeitung ist das Fehlen eines Scharfzeichnungsfilters.

PixEdit 4.0 (Techsoft)

Das Bildverarbeitungsprogamm PixEdit ist für die Bearbeitung von bis zu DIN A0 großen bitonalen Rasterdateien geeignet. Das Programm erlaubt photorealistisches Zoomen und unterstützt die Ebenentechnik (Layer-Technik). Es ermöglicht im Batch-Betrieb über Filterfunktionen eine Reduktion oder ein Entfernen von Störpixeln und eignet sich somit zum *Image Enhance-*

ment von 1 Bit-Images. Dabei kann jeder Filter so konfiguriert werden, daß das Filterergebnis gelöscht oder auf einen anderen Layer verschoben wird.

Bei der Bearbeitung von 1 Bit Images mit PixEdit hat sich gezeigt, daß mittels der Filterfunktion sehr wohl eine visuelle Verbesserung der Bildqualität erreicht werden kann, indem Bildstörungen beseitigt werden. Allerdings werden durch den Filtereinsatz oft auch Bildinformationen entfernt, die z.B. für eine korrekte OCR-Erfassung notwendig sind, so daß eine Nachbereitung von Images für eine OCR-Erkennung mit PixEdit nicht sinnvoll erscheint.[3]

ACDSee 32 2.3/2.41 (ACD Systems Ltd.)

Das Programm wurde neben ThumbsPlus als weiteres Hilfsprogramm für die visuelle Qualitätskontrolle eingesetzt. Es handelt sich hierbei um einen schnellen Image-Viewer mit Slideshow-Modul und genauer Anzeige der Dateimerkmale. Die Software ist auch für größere Dateien geeignet. Außerdem sind Batch-Konvertierungen in hoher Geschwindigkeit ins JPEG-Format möglich.

Xnview V1.09De (Pierre-e Gouglelet)

Der Freeware-Viewer unterstützt alle relevanten Bildformate und ist für mehrere Plattformen (u.a. auch Unix) erhältlich. Das Programm zeichnet sich vor allem durch einfache Bedienung und eine hohe Performance aus.

Internet Explorer 4.0/5.0 (Microsoft)

Der Web-Browser ist der bisher technisch am weitesten entwickelte. Er weist gegenüber der Konkurrenz auch bei der Verwendung von JavaScript die höchste Stabilität auf.

Netscape Navigator/Communicator 4.0x/4.5x/4.6 (Netscape)

Im Vergleich zum Microsoft-Web-Browser verursacht der Netscape-Browser mehr Probleme bei der Erstellung von Präsentationsmodellen und bei der Verwendung von JavaScript und DHTML. Er ist jedoch ähnlich weitverbreitet wie der Internet Explorer und daher nicht zu vernachlässigen.

---

[3] Siehe dazu in diesem Band den Beitrag von Thomas *Fricke* und Gerald *Maier:* Automatische Texterkennung bei digitalisiertem Archiv- und Bibliotheksgut. S. 201.

HTML-Editoren

HTML EDITOR Phase 5 (Ulli Meybohm)

Das Freeware-Programm wurde als Editor für die Entwicklung von Internetpräsentationen verwendet, da dieser Editor einen sehr großen Funktionsumfang besitzt und in seiner codebasierten Ausrichtung für professionelles Webdesign geeignet ist.

Frontpage 98 (Microsoft)

Das kommerzielle Webdesign-Paket wurde in erster Linie für die Entwicklung der Buchmodelle eingesetzt.[4] Mit dem Programm können sog. Wizzards erstellt werden, die eine benutzerfreundliche Administrierung und geführte Erstellung von Buchmodellen ermöglicht.

Automatische Texterkennung (OCR)

FineReader 4.0a Standard/Handprint (ABBYY)

Die Texterkennungssoftware FineReader wurde für die verschiedenen OCR-Tests eingesetzt. Getestet wurde dabei die Standard-Version und die Handprint-Version. Das Programm bietet eine für Standardsoftware sehr hohe Erkennungsrate und ist einfach zu bedienen. Das Training von Zeichen wird unterstützt, ebenso eine automatische und manuelle Segmentierung von Tabellen. Die Texterkennung im Stapelverfahren ist möglich. Eine Rechtschreibprüfung ist eingebaut, bei der neue Wörter zum Wörterbuch in verschiedenen Sprachen hinzugefügt werden können. Unterstützt werden Grafikdateien im 1 Bit, 8 Bit und 24 Bit-Modus.

Gegenüber der Standardversion bietet die Handprint-Version eine Erkennungsfunktion von handgeschriebenen Blockschrift-Buchstaben und eine automatische Überprüfung der Erkennungsergebnisse durch reguläre Ausdrücke und Prüfregeln. Außerdem wird ein automatischer Ausgleich von Scannerverzerrungen unterstützt.

---

[4] Siehe dazu in diesem Band den Beitrag von August *Wegmann:* Elektronische Bücher im Internet. S. 287.

Omnipage Professional 8.0 (Caere)

Neben FineReader wurde das Programm OmniPage Professional für den OCR-Feldtest eingesetzt. Das Programm arbeitet bei der Zeichenerkennung nicht mit dem Vergleich von vorgefundenen und trainierten Zeichen-Formen (pattern matching), sondern mit der Analyse von Eigenschaften der vorgefundenen Zeichen (feature recognition). Dies hat den Vorteil, daß auf eine Trainingsphase verzichtet werden kann. Dadurch ist die Bedienung des Systems recht einfach. Das Trainieren einzelner Zeichen ist aber möglich. Die Erkennungsrate ist insgesamt etwas niedriger als beim Konkurrenzprodukt FineReader.

CD-R-Software

WinOnCD 3.5/3.6 (CeQuadrat)

Das CD-R-Brennprogramm wurde in beiden Versionen getestet und hat sich als stabile Brennsoftware mit komfortabler Bedienung erwiesen.

Nero 4.0 (Ahead)

Das CD-R-Brennprogramm wurde in der im Internet erhältlichen 30 Tage-Testversion getestet. Es bietet ein schnelles, komfortables Brennen von CD-Rs. Das Programm ist ausgereift und läuft stabil.

CD-R Diagnosis V1.4.x (CD-ROM Production)

Das Programm liefert u.a. Informationen über den Inhalt, Dateiformat, physische Beschichtung und Lesbarkeit bzw. Fehler von CD-ROMs, CD-Rs und CD-RWs. Es wurde für die Kontrolle von CD-R-Datenträgern verwendet. Für wiederbeschreibbare CD-RW-Disks besitzt die Software auch eine Funktion zur Wiederherstellung beschädigter Daten.

Sonstige Software-Tools

Silverfast Ai IT–8 V4.x Scan Software (Lasersoft)

Für Testscans mit dem Flachtbettscanner und Diascanner wurde als Scansoftware Silverfast Ai IT–8 V4 eingesetzt. Die Software bietet eine einheitliche Benutzeroberfläche für verschiedene Scangeräte. Es werden verschiedene Einstellmöglichkeiten geboten, um schon während des Scanvorgangs optimale Scanergebnisse zu erzielen. Von besonderer Bedeutung ist die Möglichkeit der Gerätekalibrierung und Profilgenerierung für den jeweiligen Scanner mittels IT-8-Testvorlagen.

Rename-it 3.1 (Kent H.)

Das Freeware-Programm ist ein mächtiges Werkzeug für die Umbenennung von Dateinamen im Batch-Betrieb. Mit dem Programm können auch sequentielle Bilddateien automatisch neu numeriert werden. Für die Erstellung von Buch-Präsentationsmodellen mit sequentiellen Image-Folgen ist es ein unerläßliches Werkzeug.

InfoRapid Suchen & Ersetzen 2.1 (Ingo Straub Softwareentwicklung)

Das für den privaten Gebrauch als Freeware erhältliche Programm wurde als Werkzeug für das Suchen und Ersetzen von Textelementen bei der Gestaltung von Internet-Präsentationsmodellen eingesetzt. Es ermöglicht eine rationelle Vereinheitlichung und Anpassung der Formatierung.

Xenu's Link Sleuth 1.0r (Tilman Hausherr)

Der Freeware-HTML-Validator wurde zur Überprüfung der Link-Struktur des Projektinternetangebots eingesetzt.

2. Verwaltung digitaler Konversionsformen

Als Ergebnis des Scannens entstehen einzelne Bilddaten. Diese müssen nach bestimmten Kriterien organisiert und verwaltet werden, damit in den folgenden Schritten die Bildqualität kontrolliert werden kann und die Bilder rationell weiterverarbeitet werden können.

Benennung der digitalen Formen des *großen und kleinen Warenkorbs*

Entscheidend für die Verwaltung digitaler Formen ist die Identifikation der einzelnen Objekte mit ihren Dokumenten. Unter Objekte werden hier einzelne Archivalien oder Bücher verstanden, deren einzelne Seiten wiederum als Dokumente aufgefaßt werden können.

Wichtig – auch aus archivfachlicher Sicht – ist dabei, daß ein Zusammenhang der einzelnen Dokumente eines Objekts erhalten bleibt, u.a. durch die Wahl entsprechender Verzeichnisstrukturen.

Die Identifikation einzelner Objekte und Seiten erfolgt über den Dateinamen der digitalen Formen. Die Benennung der Dateinamen wurde dabei analog zur Kennung der Filmaufnahmen vorgenommen. Für den *kleinen*

Warenkorb wurde analog zum großen ebenfalls ein System von medienübergreifenden Kennungen eingeführt.[5]

Die Kennzeichnung der CD-R-Datenträger, auf denen die digitalen Formen von den Dienstleistern geliefert wurden, erfolgte durch die Dienstleister. Der Zugriff auf einzelne Warenkorb-Nummern bzw. einzelne Dokumente einer Nummer wurde über eine Konkordanz ermöglicht, die als Datenbank-Datei im Microsoft Access-Format vom Dienstleister erstellt worden war und als Ausgangsbasis für eine elektronische Weiterverarbeitung diente.

Verwaltung der Bilder mit Bildverwaltungssoftware

Zur Verwaltung der Objekte und der einzelnen Dokumente einschließlich der visuellen Identifikation wurde für die Projektarbeit das Bildverwaltungsprogramm ThumbsPlus (Cerious Software) eingesetzt.

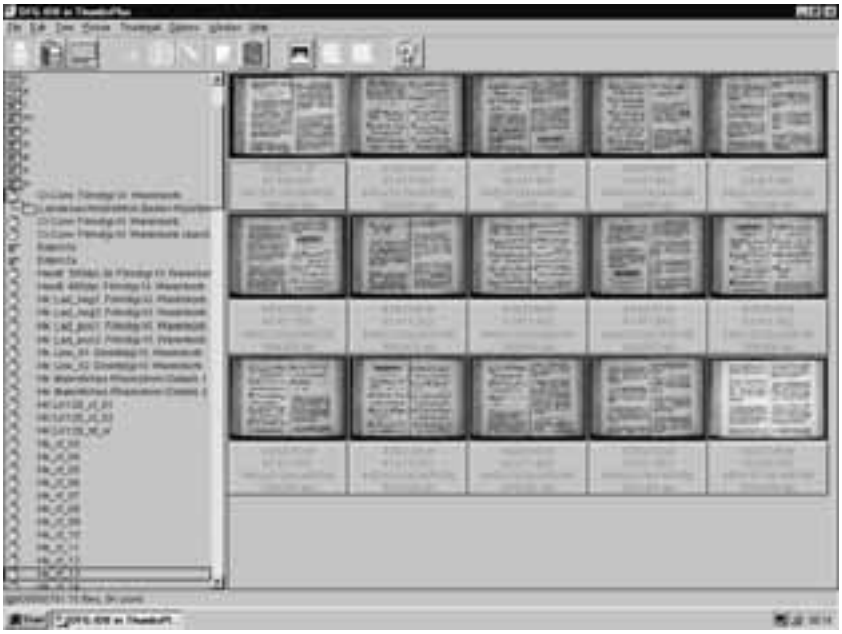


Abb. 1: ThumbsPlus 3.30: Verzeichnisbaum und Thumbnailgenerierung: Bildverwaltung, Vollständigkeitskontrolle und Kontrolle der Lesbarkeit.

---

5 Siehe dazu in diesem Band den Beitrag von Peter *Exner:* Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut. S. 113.

Dieses Programm ermöglicht, die einzelnen CD-Rs einschließlich der gespeicherten Bilder in einer Bilddatenbank zu erfassen, die File-Strukturen abbildet. Erfaßt werden dabei

- die Datenträgerbezeichnung der jeweiligen CD-R,
- die Verzeichnisstruktur einer CD-R,
- und die einzelnen Bild-Dokumente.

Außerdem wurde mit der Software für jedes Bild ein sogenannter Thumbnail, d.h. ein Vorschaubild, generiert, der ebenfalls in der Bilddatenbank gespeichert wurde. Die Thumbnails dienen der visuellen Identifikation einzelner Objekte und Dokumente.

3. Qualität digitaler Konversionsformen

Die Verwaltung der Bilder ist Voraussetzung für eine Kontrolle der Bildqualität und für ihre weitere Verwendung und Verarbeitung bis hin zur Präsentation.

Das Thema *Qualität* wurde unter folgenden Aspekten untersucht:

- Qualitätssicherung
- Kontrolle der digitalen Formen auf Vollständigkeit und Lesbarkeit
- Kriterien für die Beurteilung der Qualität von digitalen Konversionsformen
- Visuelle Kontrolle der Bildqualität, verbunden mit einem Qualitätsvergleich *Direktdigitalisierung vs. Filmdigitalisierung*
- Colormanagement als eine Voraussetzung für Bildqualität
- Ein Vergleich verschiedener Dateiformate (Abbildungsqualität, Komprimierung, Dateigröße)

Qualitätssicherung

Um eine gleichbleibende Bildqualität zu erreichen, müssen Maßnahmen zur Qualitätssicherung vorgenommen werden. Eine solche wird erreicht durch

- visuelle Qualitätskontrolle,
- die Berücksichtigung von Faktoren, die die Bildqualität beeinflussen
- und die Anwendung erprobter, normgerechter Bearbeitungsverfahren (z.B. richtige Reihenfolge verschiedener Bearbeitungsschritte in einem Workflow von der Mikroverfilmung über die Filmdigitalisierung bis hin zur Bildbearbeitung; Einsatz von Colormanagement).

Die Bildqualität beeinflussende Faktoren

Für eine Qualitätssicherung ist Voraussetzung, daß die Faktoren, welche die Bildqualität beeinflussen, bekannt sind. Folgende Faktoren sind für die Qualität digitaler Bilder von Bedeutung:

- die Qualität und Vorbereitung der Scanvorlagen, also auch der Scanfilme (u.a. richtige Positionierung der Objekte, passender Hintergrund),[6]
- die Leistungsfähigkeit der verwendeten Ein- und Ausgabegeräte,
- die Leistungsfähigkeit der verwendeten Software (Scansoftware, Viewer, Bildbearbeitungssoftware),
- die verwendete Eingabeauflösung,
- die gewählte Farbtiefe und der Tonwertumfang,
- der gesetzte Schwellenwert (bei 1 Bit-Digitalisierung),
- effektives Colormanagement inklusive Gerätekalibrierung,
- das verwendete Dateiformat und der verwendete Komprimierungsfaktor,
- erfolgte Bildnachbearbeitung (Image Enhancement) und
- Fertigkeiten und Qualifikation des Scandienstleisters

Richtige Vorbereitung der Scanvorlage für die Digitalisierung

Von den genannten Faktoren, die die Bildqualität beeinflussen, ist die richtige Vorbereitung der Vorlage für die Digitalisierung von besonderer Bedeutung. So darf das Dokument nicht gedreht sein. Die Zeilen sollten waagerecht liegen. Im Zweifelsfall ist die Orientierung an den Rändern des Trägermediums (Papiers) möglich. Besonders wichtig bei Serien (z.B. Seiten eines Buches, Akten mit gleichem Format oder Teilen einer Karte usw.) ist das Ausrichten aller Objekte an einem bestimmten Fixpunkt. Auch bei der Erstellung von Scanfilmen sollte darauf geachtet werden, daß die Aufnahme der Teile nicht mit unterschiedlichem Neigungswinkel geschieht. Durch das Drehen des eingescannten Bildes entstehen Qualitätsverluste. Das Herausfinden des geeigneten Drehwinkels ist eine aufwendige Zusatzarbeit. Bei Scanfilmen für die Filmdigitalisierung sollten wechselnde Verkleinerungsverhältnisse und eine wechselnde Ausrichtung der Dokumente auf dem Film vermieden werden. Auch breite Ränder um das Objekt sind zu vermeiden (unnötiger Speicher-

---

[6] Insgesamt ist die Qualität eines digitalen Bildes begrenzt durch die Qualität der Scanvorlage. Zur objektgerechten Mikroverfilmung und den verschiedenen Mikroformen siehe auch Hartmut *Weber:* Verfilmen oder Instandsetzen? Schutz- und Ersatzverfilmung im Dienste der Bestandserhaltung. In: Bestandserhaltung in Archiven und Bibliotheken. Hg. von Hartmut *Weber* (= Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg A 2). Stuttgart 1992. S. 103–111.

platzverbrauch, verlangsamte Verarbeitung). Damit das Objekt problemlos freigestellt werden kann, ist ein homogener bzw. möglichst monochromer Hintergrund zu wählen, der sich im Tonwert gut vom Dokument absetzt. Für die Filmdigitalisierung von gleichförmigen Objekten (z.B. Buchdoppelseiten) mit Mikrofilmscannern sollte auf dem Mikrofilm eine für den Filmschritt auswertbare Bildmarke (*Blip*) angebracht sein, die eine exakte, automatische Positionierung des Scanfensters ermöglicht.

Wahl einer objektgerechten Farbtiefe und Scanauflösung

Im Vorfeld des Digitalisierens ist neben der richtigen Vorbereitung der Scanvorlage auch die Wahl der objektgerechten Farbtiefe (1 Bit, 8 Bit oder 24 Bit) und Scanauflösung (Eingabeauflösung in dpi) entscheidend. Diese Parameter sollten in Abhängigkeit von Beschaffenheit und Homogenität der einzelnen Objekte und Dokumente (etwa nur Schrift, Handschrift oder Maschinenschrift, ab und zu fotografische Bilder oder Rasterbilder, unterschiedliche Formate usw.) gewählt werden.[7]

Um hier nähere Erkenntnisse zu bekommen, wurde der *kleine Warenkorb* mit einer Farbtiefe von 1 Bit und unterschiedlichen Eingabeauflösungen (400–800 dpi) mit einem Flachbettscanner digitalisiert, anschließend wurden die einzelnen Warenkorb-Objekte auf eine Eignung für die 1 Bit-Digitalisierung hin bewertet. In diesem Rahmen wurde auch der für die Qualität der 1 Bit-Digitalisierung besonders wichtige Faktor *Schwellenwert* durch verschiedene Einstellungen getestet.[8] Unter verschiedenen Schwellenwert-Einstellungen wurde der jeweils objektgerechte Wert ermittelt.

Dabei hat sich gezeigt, daß der passende Schwellenwert in der Regel mit Hilfe der Scanvorschau oder eines Testscans weitgehend richtig eingestellt werden kann. Die optimalen Schwellenwerte schwanken dabei stark zwischen 50–150 innerhalb einer Skala von 0–255. Erwartungsgemäß wurde mit einer Eingabeauflösung von 600 dpi eine bessere Glättung der Konturen bei Schriften erreicht. Einzelne Tests mit 800 dpi ergaben keine zusätzlichen sichtbaren Verbesserungen.

Druckschriften, Schreibmaschinenschriften und Strichgrafiken eignen sich gut für die bitonale Digitalisierung. Dies gilt natürlich auch für die Filmdigitalisierung, wenn solche Vorlagen auf AHU-Mikrofilm verfilmt

---

7 Vgl. dazu Marianne *Dörr* und Hartmut *Weber:* Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft. In: Zeitschrift für Bibliothekswesen und Bibliographie 44 (1997), Heft 1, S. 64f. und online unter: *http://www.lad-bw.de/lad/dfgdigh1.htm.*

8 Der sog. Schwellenwert beim Scannen im 1 Bit-Modus bestimmt, welche Pixel in schwarz und welche in weiß umgesetzt werden.

sind.[9] Wenn der Schwellenwert richtig eingestellt ist, können selbst blasse Handschriften auf dunklem Papier reproduziert werden. Nicht geeignet für die 1 Bit-Digitalisierung sind Kohlepapier-Durchschriften bei Schreibmaschinentexten, beidseitig beschriebenes Pergaminpapier oder Schriftstücke, die mit unterschiedlichen Schreibmaterialien und Farben beschrieben sind. Dies gilt besonders für starke Kontrastabweichungen auf dunklem Papier wie heller Bleistift kombiniert mit Tinte. Außerdem werden Objekte mit extrem feinen und schwach getönten Linien im 1 Bit-Modus schlecht wiedergegeben. Erwartungsgemäß wurden Halbton-, Graustufen- und Farbbilder in diesem Modus nur sehr schlecht reproduziert, da Teile der Bildinformation und Farben völlig verloren gehen.

So sind Farbtiefen von 8 Bit (Graustufen) bzw. 24 Bit (Farben) normalerweise für die Digitalisierung von Bleistiftzeichnungen, Illustrationen, Photographien, Halbton- und Farbmikrofilmen erforderlich. Dabei gilt, daß die Wahl einer höheren Farbtiefe eine Reduktion der Scanauflösung zuläßt. Als sinnvolle Scanauflösungen auch im Hinblick auf die zu erwartenden Dateigrößen haben sich bei Farbtiefen ab 8 Bit für digitale Masterformen 300–400 dpi erwiesen.

Kontrolle der Datenträger und der digitalen Formen auf Vollständigkeit und Lesbarkeit

Da im Rahmen des Projekts die Digitalisierung fast ausschließlich durch externe Dienstleister erfolgte, waren neben einer Bewertung der Bildqualität auch Verfahren zur Kontrolle der Datenträger auf Vollständigkeit und Lesbarkeit der einzelnen Bilddateien erforderlich.

Zunächst mußte dabei die Lesbarkeit und Fehlerfreiheit der Datenträger selbst geprüft werden. Sie erfolgte für CD-Rs mittels der Spezialsoftware CD-R Diagnosis 1.4.3 (CD-ROM Productions). Das Programm liefert u.a. Informationen über Dateiformat, physische Beschichtung und Lesbarkeit bzw. Fehler von CD-Rs.

In einem nächsten Schritt wurde im Rahmen der Vollständigkeitskontrolle die korrekte Dateinamenbezeichnung der Bilder geprüft. Die Vollständigkeitskontrolle erfolgte zu einem großen Teil ebenfalls über das schon vorgestellte Programm ThumbsPlus. So wurde mit der schon erwähnten Thumb-

---

[9] AHU (Anti-Halation-Undercoated)-Mikrofilme sind mit einer sogenannten Lichthofschutzschicht versehen, wodurch die Schärfe der photographischen Abbildung verbessert wird. AHU-Filme sind panchromatisch, kontraststeigernd und eignen sich vorzugsweise für Strichvorlagen. Vgl. dazu Barbara *Keimer:* Mikroverfilmung von Büchern als bestandserhaltende Maßnahme (= dbi Materialien 154). Berlin 1997. S. 15.

nail-Generierung zugleich die Lesbarkeit einzelner Bild-Dateien geprüft. Außerdem ermöglicht die Software durch die Strukturierung der Bilder eine gute Kontrolle auf Vollständigkeit, die vor allem bei Objekten mit sequentiellen Dokumenten geprüft werden mußte.

Neben der Lesbarkeit der Bild-Dateien sollten auch sichtbare Bilddefekte erkannt werden wie z.B. halbe, abgeschnittene Bilder. Dafür diente als weiteres Hilfsprogramm die Viewer-Software ACDSee 32 2.x.

Für Stichproben wurde auch die Bildbearbeitungssoftware Adobe Photoshop eingesetzt, die wie die vorgenannte Viewer-Software neben einer 1:1 Darstellung auch verschiedene Zoom-Stufen für die Qualitätskontrolle und Werkzeuge für die Bildbearbeitung bietet.

Kriterien für die Beurteilung der Qualität von digitalen Formen und die visuelle Qualitätskontrolle

Vom Projektteam wurden im Rahmen der visuellen Qualitätskontrolle und der theoretischen Beschäftigung mit dem Thema *Bildqualität* folgende Hauptkriterien für die Qualität digitaler Konversionsformen festgelegt:

1. Bildschärfe:
   Ziel ist die Wiedergabe des kleinsten abzubildenden Teils eines Bildes, die Wiedergabe scharfer Konturen und Lesbarkeit von Schriften.
2. Tonwiedergabe:
   Ziel sind Helligkeitsabstufungen, die möglichst denen der Originalvorlage entsprechen (Einfluß der Farbtiefe).
3. Farbwiedergabe:
   Ziel ist eine ausgewogene Farbwiedergabe, die derjenigen der Originalvorlage entspricht.
4. Mögliche Bildstörungen:
   Ziel ist die Vermeidung ungewollter, zufälliger Bildstörungen (z.B. Rauschen, Moiré, Streupixel).

Die visuelle Kontrolle der Bildqualität konzentrierte sich insbesondere auf die beiden Kriterien Bildschärfe und Tonwiedergabe.

Die Bildschärfe als Qualitätskriterium und ihre Beurteilung

Die Bildschärfe ist das hauptsächliche Qualitätsmerkmal bei bitonalen (1 Bit)-Scans, also bei der Digitalisierung von maschinenschriftlichen bzw. gedruckten Textvorlagen und Strichzeichnungen. Sie ist vor allem von der Wahl der geeigneten Scanauflösung abhängig, aber auch von einer korrekten Focusierung beim Scannen.

Als Kontrollmedien für die Beurteilung der Bildschärfe dienten ein hochauflösender 21"-Monitor mit Lochmaske (0,27 mm)[10] in einer Bildschirmauflösung von 1280 x 1024, 24 Bit Farbtiefe, und die schon erwähnte Viewer Software ACDSee 32 V2.41, die eine visuelle Qualitätskontrolle in verschiedenen Vergrößerungsstufen ermöglicht. Bei bitonalen Bildern erfolgt die Kontrolle zusätzlich über einen mit Laserdrucker erstellten Papierausdruck.

Die Beurteilung der Bildschärfe erfolgte über Testtafeln mit Testzeichen, Testbildern und durch das Testen verschiedener Schriftgrößen auf Lesbarkeit. Dabei wurde auch überprüft, ob der vorgegebene QI (Quality Index)-Wert erreicht wurde.[11]

Zunächst wurde die im *kleinen Warenkorb* enthaltene Testtafel *Image Evaluation Test Target (MT-2)* in filmdigitalisierter Form (1 Bit und 8 Bit mit jeweils 600 dpi) visuell beurteilt nach der Lesbarkeit bestimmter Schriftgrade der Schrift *Futura News Gothic.* Anschließend wurde der entsprechende QI bestimmt.

Der Test zeigte, daß bei einer Eingabeauflösung von 600 dpi sowohl bei der Direktdigitalisierung als auch bei der Filmdigitalisierung im 1 Bit-Modus und im 8 Bit-Modus QI 8, d.h. hohe Qualität, erreicht werden kann. Dabei können sowohl bei der Direktdigitalisierung als auch bei der Filmdigitalisierung ein Schriftgrad von 6 Punkt bei einem 1 Bit-Scan noch gut gelesen werden. Das Scannen mit Graustufen (8 Bit) bringt bei gleicher Eingabeauflösung bei beiden Digitalisierungsverfahren eine Qualitätssteigerung der Lesbarkeit auf bis zu 4 Punkt.

Weitere Tests für Bildschärfe und Auflösung mit den ISO-Testzeichen Nr. 1 und 2[12] der Testtafeln T2 und T3 und Schriftproben der Testtafel T3[13] des *kleinen Warenkorbs* bestätigten das Ergebnis.

---

10 IIyama Vision Master 500 (MF–8721).

11 Zur Bedeutung und Anwendung des Quality Indexes (QI) für die Beurteilung der Bildqualität von digitalen Formen siehe: Digital Imaging for Libraries and Archives. Ed. by Anne R. *Kenney* and Stephen *Chapmann.* Cornell University Library. New York 1996. P. 12–31. Dabei bedeutet QI 8= *hohe Qualität*, QI 5= *mittlere Qualität* und QI 3= *niedere Qualität.*

12 Zum Gebrauch der ISO-Testzeichen Nr. 1 und 2 siehe DIN-Taschenbuch 154. Publikation und Dokumentation 2. Normen. Hg. vom DIN-Institut für Normung e.V. Berlin u.a. ^{4}1996. S. 76–84. Das ISO-Testzeichen Nr. 1 hat die Form eines gleichseitigen Achtecks (Oktogon), in dem zwei Linien parallel zu einer Seite des Achtecks angebracht sind. Die Größe des Testzeichens wird durch die Zeichenbreite ausgedrückt, das Maß für die Zeichenbreite (Schlüsselweite) wird Kennzahl genannt und bei Testtafeln in 1/100 angegeben. ISO-Testzeichen Nr. 2 (siehe Abb. 2) besteht aus zwei zueinander senkrecht stehenden Liniengruppen aus je fünf dunklen und fünf hellen Linien (Linienpaare). Die Kennzahl gibt die dargestellten Linienpaare je mm an.

13 Die Testtafel enthält Schriftproben mit und ohne Serifen der Schriftgrößen 4 Punkt bis 14 Punkt in 2-Punkt-Schritten.

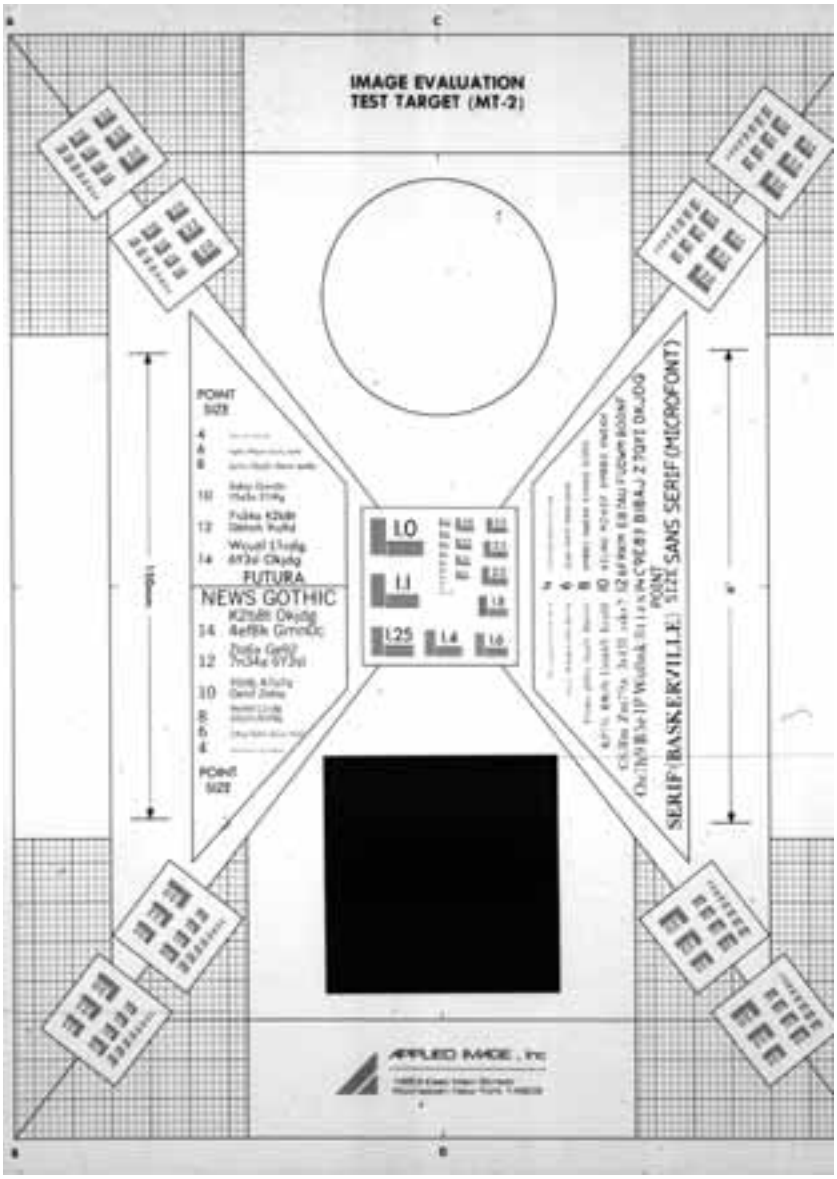


Abb. 2: Image Evaluation Test Target (MT-2) mit ISO-Testzeichen Nr. 2 (*kleiner Warenkorb,* T3).

Nr.	Ausgangsform	T2: ISO-Testzeichen Nr. 1 (Kennzahl)	T3: ISO-Testzeichen Nr. 2 (Lp/mm)	T3: Schriftgr. *Serif Baskerville* (in Punkt)[14]	T3: Schriftgr. *Futura News Gothic*, Sans Serif (in Punkt)
1	Vorlagen	35	18	4 (8)	4 (8)
2	Mikrofilm, AHU	35	18	4 (8)	4 (8)
3	Mikrofilm, Halbton	35	18	4 (8)4	4 (8)
4	Diafilm „Kodak Ektachrome“	70	5,6	4 (5)	4 (5)
5	Zeutschel OmniScan Aufsichtscanner, 600 dpi, 1 Bit	–	5	6 (5) 8 (8)	4 (3) 6 (5)
6	Zeutschel OmniScan Aufsichtscanner, 600 dpi, 8 Bit	42	8	4 (5)	4 (8)
7	Sunrise ProScan IV-Filmscanner, 600 dpi, 1 Bit, von AHU-Film	84	4,5	6 (5) 8 (8)	6 (8)
8	Sunrise ProScan IV-Filmscanner 600 dpi, 8 Bit, von Halbtonfilm	60	5,6	4 (3) 6 (8)	4 (3) 6 (8)

So sind Auflösung und Bildschärfe entsprechend der Testzeichen- und Schriftgrößenbeurteilung bei der Direktdigitalisierung mit Aufsichtscanner und der Filmdigitalisierung beinahe identisch, bei jeweils gleicher Eingabeauflösung (vgl. Nr. 5 mit Nr. 7). Auch bei diesem Vergleichstest zeigte sich, daß das Scannen mit Graustufen (8 Bit) bei beiden Digitalisierungsverfahren bei gleicher Eingabeauflösung zu einer besseren Lesbarkeit führt. Vergleicht man die digitalisierten Formen mit der Originalvorlage und der Mikroverfilmung, dann wird deutlich, daß selbst bei einer Digitalisierung mit 800 dpi in 8 Bit Farbtiefe ein Mikrofilm mehr als doppelt soviel Linienpaare pro mm bei ISO-Testzeichen Nr. 2 auflösen kann. Der Mikrofilm ist der Originalvorlage qualitativ ebenbürtig (vgl. Nr. 1 mit Nr. 2 und 3) und bietet ausreichende Reserven als Ausgangsbasis für die Digitalisierung.

[14] Die Zahlen in ( ) stehen für den entsprechenden Quality-Index (QI)-Wert.

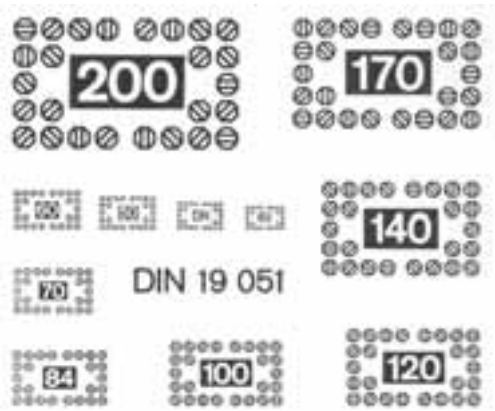


Abb. 3: ISO-Testzeichen Nr. 1.

Die Tonwiedergabe als Qualitätskriterium und ihre Beurteilung

Neben der Bildschärfe ist die Tonwiedergabe ein zweites wichtiges Qualitätskriterium. Die Tonwiedergabe, d.h. der Grad der Wiedergabe der im Original vorhandenen Helligkeitsabstufungen, ist schwierig zu kontrollieren, da alle Komponenten eines Digitalisierungs-Systems Einfluß auf sie haben.

Die korrekte Tonwiedergabe ist vor allem ein wichtiges Qualitätskriterium für Graustufen- und Farbscans. Bei 1 Bit-Scans mit den beiden Tonwertstufen schwarz und weiß ist der schon genannte Schwellenwert für eine korrekte Wiedergabe entscheidend.

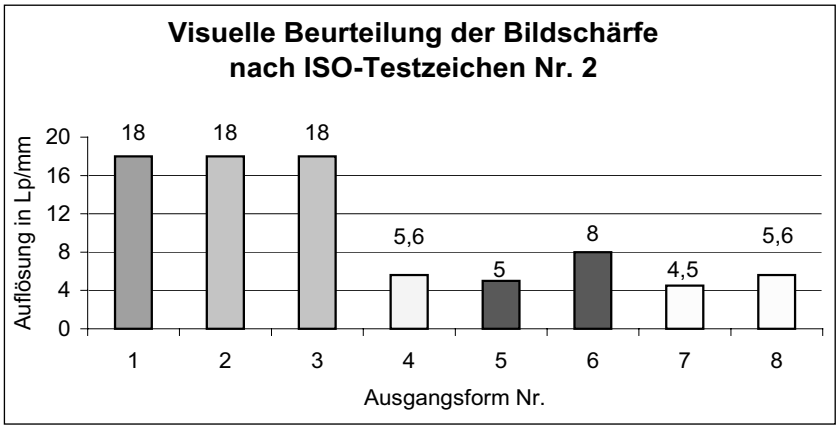


Abb. 4: Visuelle Beurteilung der Auflösung nach ISO-Testzeichen Nr. 2 bei verschiedenen Ausgangsformen.

Bei einem gescannten Bild muß im Vergleich zur Scanvorlage hinsichtlich der Tonwiedergabe folgendes geprüft werden:

- die gleichmäßige Detailabbildung in hellen und dunklen Stellen des Bildes,
- ein gleichmäßiger Gradationsverlauf,
- die Existenz von Moiré-Mustern und anderen Störbildern und
- die Existenz signifikanter im Original enthaltener feiner Details.

Für eine erste Überprüfung von Helligkeit und Tonwertabstufungen bei sequentiellen digitalisierten Bilderserien wurde die Thumbnail-Funktion der Software ThumbsPlus verwendet. So sollten Abstufungen und Helligkeit bei in sequentieller Form vorliegenden Objekten, wie Büchern oder mehrseitigen Akten, innerhalb eines Objektes gleichmäßig sein. *Ausreißer* konnten somit schnell erkannt werden. Die genauere Überprüfung erfolgte mit Hilfe des Histogramms unter Verwendung der Bildbearbeitungssoftware Photoshop.

Die Bedeutung des korrekten Tonwertumfangs mit einem Vergleich von Direkt- und Filmdigitalisierung kann an einem Internet-Testmodell mit einer farbigen Wappendarstellung gezeigt werden.[15]

Hier können die verschiedenen Formen (Filmdigitalisierung, Direktdigitalisierung, unbearbeitet, bearbeitet) anhand einer *Text*-Seite und einer *Bild*-Seite des Wappenbuchs miteinander verglichen werden:

D0:F0 Die Ausgangsformen weisen bedingt durch die Aufnahmetechnik unterschiedliche Tonwerte auf:
- Der verwendete Farbmikrofilm ist kontraststeigernd, daher ist der Kontrast bei der Filmdigitalisierung eher zu hoch, die dunklen Bereiche sind zu wenig differenziert. Eine partielle Aufhellung der dunklen Stellen und eine Kontrastdämpfung sind notwendig.
- Bei der Direktdigitalisierung ist das Ergebnis eher flau, wirkt aber in den Tönen differenzierter. Eine Kontraststeigerung durch Nachbearbeitung ist notwendig.

D1:F1 Zusätzlich weisen beide Formen Unschärfen auf, die über Schärfen-Filter ausgeglichen werden können.

D2:F2 Das Ergebnis nach der Tonwertkorrektur über Gradationskurve und Histogramm zeigt, daß eine weitgehende Angleichung der Qualität durch Bildverbesserung möglich ist.

Insgesamt zeigen die Masterformen zunächst deutliche Unterschiede in der Abstufung der Helligkeitsverteilung und der Abstufung der Farbtöne. Die

---

[15] Siehe Internet-Testmodell *Ehevertrag Eberhards IV. von Württemberg mit Henriette von Mömpelgard mit Wappen der Häuser Württemberg und Mömpelgard* (aus: Hauptstaatsarchiv Stuttgart, A266, U1, Ende 14. Jahrhundert) unter *http://www.lad-bw.de/digpro/graf.htm.*

Abb. 5: Testmodell *Wappen der Häuser Württemberg und Mömpelgard (http://www.lad-bw.de/digpro/graf.htm).*

filmdigitalisierte Form bietet in diesem Fall zunächst eine schlechtere Darstellungsqualität, die sich auch in einem ungünstigen Verlauf der Gradationskurve und einer ungleichmäßigen Helligkeitsverteilung des Histogramms manifestiert. Die Darstellungsqualität konnte aber durch Maßnahmen der Bildverbesserung, d.h. durch eine Tonwertkorrektur und in diesem Fall Kontrastdämpfung verbessert und der direktdigitalisierten Form angeglichen werden.

Bewertung der Qualität von filmdigitalisierten 1 Bit und 8 Bit Bildern im Rahmen der visuellen Qualitätskontrolle

Die von den Dienstleistern gelieferten filmdigitalisierten Master weisen teilweise eine schlechtere Qualität auf als ihre direktdigitalisierten Pendants. Dies ist nicht auf die Qualität der Scanfilme zurückzuführen, die vor allem in Hinblick auf das Auflösungsvermögen hohe Qualitätsreserven besitzen, sondern zu einem großen Teil auf Einstellungsfehler beim Scannen wie eine unzureichende Fokusierung, eine ungeeignete automatische Kontrast- und Schwellenwerteinstellung beim Scannen im 1 Bit-Modus oder eine fehlerhafte bzw. unterlassene Tonwertkorrektur im 8 Bit Modus.

Bei der Digitalisierung im 1 Bit-Modus haben Einstellungsfehler, wie ein falscher Schwellenwert, wesentlich gravierendere Folgen als bei Graustufenscans. Tests mit den Objekten des *kleinen Warenkorbs* mit unterschiedlichen Filmscantechniken haben gezeigt, daß bitonale Scans, die durch eine Umwandlung von Graustufenscans (8 Bit) erstellt werden, qualitativ schlechter sind als solche, die gleich im 1 Bit-Modus erfolgen.

Bei einer zukünftigen Auftragsvergabe für Filmdigitalisierungen sollten daher zunächst Tests im Hinblick auf Bildschärfe, geeignete Schwellenwerte und Tonwerteinstellungen getroffen werden. So kann verhindert werden, daß bei fehlerhaften Schwellenwerten Haarlinien von Buchstaben ausfallen oder Linien zusammenfließen oder bei unzureichender Fokusierung trotz ausreichender Eingabeauflösung Unschärfen auftreten, die sich auch durch eine Nachbearbeitung mit Filtern nicht korrigieren lassen.

Colormanagement als Faktor für die Bildqualität

Die visuelle Qualitätskontrolle der von Dienstleistern erstellten digitalen Formen hat deutlich gemacht, daß für eine korrekte, der Vorlage entsprechenden Ton- und Farbwiedergabe, die Einführung eines sog. Colormanagements in Verbindung mit einer Charakterisierung und Kalibrierung der Ein- und Ausgabegeräte eine unabdingbare Voraussetzung ist. Dies gilt vor allem für die Farbdigitalisierung von sequentiellen Dokumenten eines Objektes oder gleichartiger Objekte.

Intention von Farbmanagement ist, Farbe und Tonwertabstufungen der Vorlage richtig in digitalisierten Formen umzusetzen. Hierbei müssen die Möglichkeiten und Grenzen der verschiedenen Farbräume beachtet werden.

Um eine kontinuierliche Qualität zu gewährleisten, ist es erforderlich, das Farbverhalten aller in die Prozeßschritte der Digitalisierung und Bildbearbeitung eingebundenen Ein- und Ausgabegeräte zu beschreiben. Hieraus ergeben sich sog. Farbprofile, die die Basis für farbverbindliches Arbeiten in den Anwendungsprogrammen bilden.

Farbmanagementverfahren wurden bisher bei einem der Dienstleister in einer Apple Macintosh Umgebung erprobt. Von seiten des Projektteams wurden theoretische Überlegungen zum Thema *Farbmanagement* angestellt und ein möglicher Farbmanagement-Workflow für die Farbdigitalisierung von Archivgut ausgearbeitet, der in einem Folgeprojekt näher untersucht werden soll.[16]

---

[16] Siehe dazu in diesem Band den Beitrag von Gerald *Maier:* Colormanagement bei der Farbdigitalisierung von Archivgut. Grundlagen, Hintergründe und Ausblick. S. 179.

4. Geeignete Dateiformate für digitale Konversionsformen (Abbildungsqualität, Komprimierung, Dateigröße)

Einfluß auf die Qualität und Dateigröße digitaler Formen hat auch die Wahl eines geeigneten Dateiformats in Verbindung mit Komprimierungstechnologien zur Reduktion der Dateigröße.

Bei der Frage nach der Verwendung eines Dateiformats muß unterschieden werden zwischen dem sog. digitalen Master und digitalen Nutzungsformen. Die Anforderungen, die an den digitalen Master gestellt werden, sind aus der Art der Digitalisierungsvorlagen abzuleiten. Auf jeden Fall sollte das Dateiformat der Archivierungsversion eine Speicherung ohne Qualitätsverluste ermöglichen.

Die digitalen Nutzungsformen werden durch Konvertierung und Bearbeitung des digitalen Masters erstellt. Sie genügen niedrigeren Qualitätsanforderungen, benötigen aber zugleich eine möglichst geringe Dateigröße, um den raschen Datentransfer in Online-Medien wie dem Internet zu ermöglichen.

Die verschiedenen Dateiformate im Vergleich

TIF (Tagged Image File)-Format

- Vorteile:
  - quasi plattformübergreifendes Standarddateiformat für digitale Master
  - unterstützt viele Farbräume u.a. RGB und LAB
  - kann ICC-Profile einbetten
  - kann Dateiinformationen im Header enthalten
  - verlustfreie Komprimierung möglich mit LZW-Algorithmus (Lempel-Ziv-Welch)
  - für 1 Bit Bilder als TIFF-G4 mit verlustfreier (Fax)-Komprimierung der Gruppe 4
- Nachteile:
  - nicht geeignet für eine Internet-Präsentation, da das Format von den aktuellen Web-Browsern nicht dargestellt und verarbeitet werden kann
  - Komprimierungsrate zu gering für Online-Präsentationen

JPEG (Joint Photographic Experts Group)-Format

- Vorteile:
  - quasi Standardformat für *digitale Nutzungsformen* im Internet
  - gute Komprimierung (Kompressionsrate zwischen 10:1 und 100:1)
  - Interlace-Effekt, 3–5 Durchgänge nach Wahl
  - Farbtiefe bis 24bit
  - verwendet die Farbtiefe des Monitors, daher entsprechende Dateigröße bei der Darstellung im Browser

- Nachteile:
  - verlustbehaftete Kompression
  - Komprimierung führt zu Bildartefakten wie verschwommenen Kanten, Blockbildung und herabgesetzter Luminanz
  - zeichnet weich, somit für Schrift, Strichgrafiken usw. eher ungeeignet (das Weichzeichnen eines Bildes ergibt einen besseren Komprimierungsfaktor; je mehr differenzierte Strukturen das Bild aufweist, desto schlechter ist der Komprimierungsfaktor)
  - Bei jedem erneuten Abspeichern einer Datei verschlechtert sich die Qualität, weil die Kompressionsfilter von neuem angewendet werden
  - unterstützt keine geräteunabhängigen Farbräume und die Einbindung von Geräteprofilen für ein Farbmanagement

GIF (Graphics Interchange Format)-Format

- Vorteile:
  - systemübergreifendes Format
  - gute Abbildungsqualität für Texte und Konturen (Bildschärfe)
  - animierbare Darstellungen, d.h. bewegte Bilder, sind über das Einfügen von Zeitsteuerungsinformationen möglich
  - Möglichkeit der Verwendung unterschiedlicher Farbpaletten
- Nachteile:
  - lizenzpflichtiger LZW-Komprimierungsalgorithmus
  - unterstützt nur eine Farbtiefe bis zu 8 Bit

PNG (Portable Network Graphics)-Format

- Vorteile:
  - plattformübergreifendes Dateiformat
  - enthält einen verlustfreien LZ77-Komprimierungsalgorithmus (verwandt mit ZIP-Packer-Format), der im Gegensatz zu demjenigen des GIF-Dateiformats nicht lizenzpflichtig ist und mit verschiedenartigen Filtern arbeitet
  - eignet sich sowohl für digitale Master als auch digitale Nutzungsformen
  - unterstützt hohe Farbtiefen (16 Bit für Graustufenbilder und sogar bis zu 48 Bit für Farbbilder, wird aber bisher von der Anwendungssoftware noch nicht unterstützt. Zum Vergleich: GIF ist auf 8-Bit begrenzt!)
  - unterstützt indizierte Farben, d.h. Farbpaletten (z.B. 256 Farben/Graustufen)
  - kann Dateiinformationen mit abspeichern
  - unterstützt linearen und schrittweisen Bildaufbau (bereits bei 20–30% der übertragenen Bilddaten ist der Bildinhalt erkennbar, bei GIF sind 50% notwendig)

  - Angaben zum Gamma-Faktor (Abweichung bezüglich des Helligkeitswertes) und zur Farbtemperatur (d.h. des Verhältnisses von Blau zu Rot) können mit abgespeichert werden (Empfänger kann sein Display entsprechend kalibrieren, so daß das Bild mit identischem Kontrast-/Helligkeitswert und Farbverlauf erscheint)
- Nachteile:
  - ältere Webbrowser unterstützen noch kein PNG-Format
  - keine so hohe Kompression wie bei JPEG oder Wavelet-Dateiformaten
  - bisher noch keine Unterstützung für geräteunabhängige Farbräume und die Einbindung von Geräteprofilen für ein Farbmanagement (Unterstützung ist aber geplant)
  - im Gegensatz zu GIF nicht animierbar

SID (MrSID-Wavelet)-Format (LizardTech)[17]
- Vorteile:
  - Verlustlose Kompression/Dekompression mit beliebiger Bildgröße und mit besseren Ergebnissen als andere verlustlose Techniken wie z.B. LZW-TIFF
  - Verlustbehaftete Kompression durch Angabe des Kompressionsfaktors und der Kompressionsqualität mit hohen Kompressionsraten mit geringen Verlusten
  - Kompression und Dekompression mit qualitativ hervorgehobenen Bildbereichen
  - Zoom-Funktion für skaliertes Laden ohne Qualitätsverlust
- Nachteile:
  - proprietärer Standard und proprietäres Dateiformat
  - im Test Probleme mit größeren Grafikdateien (>50 MB)
  - benötigt spezielle Server-Software für die Internet-Präsentation eines Bildarchivs
  - Kopierschutzmechanismen, die flexibles Arbeiten behindern
  - sehr teuer in der Anschaffung

LWF (LuraWave)- und LDF-(LuraDocument)-Format (LuRaTech Gesellschaft für Luft- und Raumfahrttechnologie)[18]
- Vorteile:
  - Verlustlose Kompression/Dekompression mit beliebiger Bildgröße und mit besseren Ergebnissen als andere verlustlose Techniken wie z.B. LZW-TIFF

---

[17] URL: *http://www.lizardtech.com.*
[18] URL: *http://www.luratech.de.*

  - Verlustbehaftete Kompression durch Angabe des Kompressionsfaktors und der Kompressionsqualität mit hohen Kompressionsraten
  - Kompression und Dekompression mit qualitativ hervorgehobenen Bildbereichen
  - Zoom-Funktion
  - durch den Benutzer steuerbares Nachladen ohne Qualitätsverlust
  - Progressives Laden mit Anzeige des Bildes während des Ladevorgangs
  - LDF: separate Kompression von Text- und Bildanteilen
  - LDF: Textanteil wird verlustlos binär kodiert
  - LDF: Text, Textfarbe – und Bildanteil sind getrennt dekodierbar
  - LDF: universelle Bereich-Maskierung möglich
  - relativ preiswert
- Nachteile:
  - proprietärer Standard und proprietäre Dateiformate
  - Bildqualität ist bei höherer Komprimierung nicht unbedingt besser als bei anderen Dateiformaten
  - LWF: benötigt spezielle Viewer-Software und Plug-ins für Internet Browser und Bildbearbeitungssoftware
  - LWF: ein gezoomtes Images läßt sich nicht bewegen bzw. scrollen
  - LDF: Textanteil wird bei Archivgut nicht immer als Text erkannt
  - LDF: bisher keine Internetanbindung möglich (ist aber in Entwicklung)

Bildqualität und Dateigröße
Ein Vergleich von Komprimierung und Farbtiefe beim PNG-, JPEG- und LWF-Dateiformat

Farbdigitalisierung (24 Bit)

Bildqualität und Dateigröße in Abhängigkeit von Komprimierung und Farbtiefe wurden an einem farbdigitalisierten Wachssiegel untersucht und in einem Internet-Testmodell visualisiert.[19] Verglichen wurden dabei die Dateiformate PNG, JPEG und LWF.

---

[19] Siehe Internet-Testmodell *Majestätssiegels Kaiser Karls IV. von 1347* (aus: Hauptstaatsarchiv Stuttgart, H51, U487) unter *http://www.lad-bw.de/digpro/graf.htm*.

Abb. 6: Testmodell *Majestätssiegel (http://www.lad-bw.de/digpro/graf.htm).*

Zunächst wurden verschiedene Komprimierungsstufen bzw. eine Farbreduktion bei PNG und JPEG untersucht. Die Ausgangsform bildete ein direktdigitalisierter, unkomprimierter 24 Bit-TIFF-Master (D0) mit eine Bildgröße von 4464 x 3300 Pixeln und einer Dateigröße von 44,2 MB. Er wurde für die Demonstration auf 424 x 439 Pixel verkleinert. Dabei hat der reduzierte Testmaster (D1) eine Dateigröße von 567,7 KB unkomprimiert (100%).[20] Im Testmodell wird der Testmaster als PNG-Datei angezeigt, da das TIF-Format in Web-Browsern nicht darstellbar ist. Die Formatkonvertierung und Bildbearbeitung erfolgte mit ThumbsPlus 3.2 und Adobe Photoshop.

Bei der Gegenüberstellung bekommt man folgendes Bild:

D2:D3 D2: Die Abspeicherung im PNG-Format mit Komprimierung, aber in voller Farbtiefe führt zu einer Reduzierung der Dateigröße mit 440,5 KB auf 77,6% der Ausgangsform.

---

20 Im Testmodell wird der Testmaster als PNG-Datei angezeigt, da das TIF-Format in Web-Browsern nicht darstellbar ist.

D3: Mit Komprimierung und Farbreduktion auf 256 Farben hat die selbe Datei im selben Format eine Dateigröße von nur noch 151,4 KB. Sie ist nun auf 26,7% der Ausgangsform reduziert.

D2:D4 D4: Das PNG-Format mit Komprimierung und Anwendung der WWW-Farbpalette (216 Farben) bringt mit 87,0 KB eine deutliche Reduktion auf 15,3% der Ausgangsgröße bei noch akzeptabler Qualität.

D2:D7 D7: Das PNG-Format mit Komprimierung und 16 Farben bringt eine weitere Reduktion der Dateigröße mit 77,0 KB auf 13,6%. Allerdings ist sie mit Verlust von Information verbunden. Dies wird deutlich sichtbar am Ausfall von Farben am mehrfarbigen Seidenfaden des Siegels.

D2:D8: D8: Die Speicherung im JPEG-Format mit hoher Komprimierung (Stufe 0) führt zu einer Reduktion auf 6,7% mit 38,0 KB. Die hohe Komprimierung bringt aber einen Qualitätsverlust in Form von Unschärfe und verringerter Sättigung mit sich.

D2:D9: D9: Die Speicherung des Objekts im JPEG-Format mit niedriger Komprimierung (Stufe 8) führt zu einer deutlichen Reduktion der Dateigröße auf 30,7% mit 174,2 KB bei geringer Bildverschlechterung (Herabsetzung der Sättigung, d.h. schlechterer Farbqualität).

Vergleichbar im Hinblick auf die Dateigröße und Bildqualität sind D3 und D9, also eine für Online-Präsentationen optimierte PNG-Datei mit einer Farbtiefe von 256 Farben (151,4 KB, Reduktion auf 26,7%) und eine JPEG-Datei mit Komprimierungsstufe 8 (174,2 KB, Reduktion auf 30,7%). Das PNG-Format bietet in diesem Fall bei gleicher oder eher besserer Bildqualität (höhere Bildschärfe) einen 4% höheren Komprimierungsfaktor. Da aber bei der Speicherung im JPEG-Format im Vorfeld keine Farbreduktion durchgeführt werden kann, behält das Bild auch im Zuge der Dekompression bei der Darstellung im Webbrowser den 24 Bit Farbmodus bei, was zu einer höheren Ressourcenbelastung des Systems führt. Bei der PNG-Version, die im Vorfeld manuell farbreduziert wurde, wird die geringe Dateigröße vorrangig von der Farbreduktion bestimmt und nicht durch den verwendeten Komprimierungsalgorithmus, so daß beim Betrachten keine starke Systembelastung stattfindet und gleichzeitig keine schärfenreduzierende Weichzeichnung erfolgt.

Das Diagramm zeigt als Ergebnis der Testreihe den Vergleich der Dateigrößen im Verhältnis zur unkomprimierten Fassung in % bei der jeweiligen Farbtiefe (PNG) bzw. Komprimierungsstufe (JPEG). Entscheidend ist dabei, daß bei entsprechender vorheriger Reduktion der Farbtiefe, bei PNG im Vergleich zu JPEG teilweise bessere Komprimierungsverhältnisse erzielt werden können, bei gleicher oder besserer Bildqualität. Dies gilt vor allem für eine auf 256 Farben reduzierte PNG-Datei (28%) im Vergleich zu einer JPEG-Datei, gespeichert in Stufe 8.

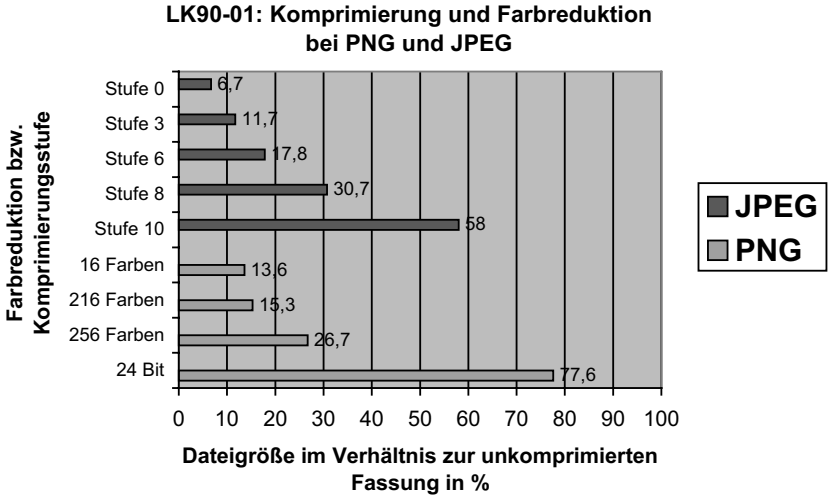


Abb. 7: Testmodell *Majestätssiegel* (Direktdigitalisierung, 24 Bit) – Komprimierung/Farbreduktion bei PNG und JPEG.

Ein ähnliches Bild zeigt sich bei einem Vergleich mit einem filmdigitalisierten 24 Bit-TIFF-Master (F0), unkomprimiert, mit eine Bildgröße von 4464 x 3300 Pixeln und einer Dateigröße von 43,2 MB. Er wurde für das Internet-Testmodell auf 424 x 451 Pixel verkleinert. Dabei hat der reduzierte Testmaster eine Dateigröße von 567,3 KB, unkomprimiert (100%).[21] Die Formatkonvertierung und Bildbearbeitung erfolgte ebenfalls mit ThumbsPlus 3.2 und Adobe Photoshop. Hier wurde nun zusätzlich das proprietäre LuraWave(LWF)-Format zum Vergleich hinzugezogen.

Im Hinblick auf die Dateigröße lassen sich hier F3, F9 und F12 miteinander vergleichen, also eine für Online-Präsentationen optimierte PNG-Datei mit einer Farbtiefe von 256 Farben (147,2 KB, Reduktion auf 26,0%), eine JPEG-Datei mit Komprimierungsstufe 8 (167,7 KB, Reduktion auf 29,6%) und eine LWF-Datei (115,0 KB, Reduktion auf 20,3%). Alle Dateien sind von vergleichbarer Qualität. Auch bei der Filmdigitalisierung bietet das PNG-Format bei gleicher oder eher besserer Bildqualität (höhere Bildschärfe) einen mehr als 4% höheren Komprimierungsfaktor im Vergleich zum JPEG-Format.

Noch besser schneidet allerdings das LWF-Format ab, das bei gleicher Bildqualität im Vergleich zu PNG eine um fast 6%, im Vergleich zu JPEG sogar eine fast 10% bessere Komprimierungsrate aufweist. Selbst bei einer noch höheren Komprimierung (F13: 1:10, 57,5 KB) auf 10,1% der Ausgangsgröße

[21] Ebd.

	Testform	KB	%
F1	Testmaster 24 Bit auf 424x451 Pixel reduziert, scharfgezeichnet, unkomprimiert	567,3	100,00
F2	24 Bit, Ton- und Farbwerte korrigiert, PNG	413,5	72,89
F3	256 Farben (Dithering, Photoshop), PNG	147,2	25,95
F4	216 Farben WWW (Dithering, Photoshop), PNG	84,3	25,95
F5	128 Farben (Dithering, ThumbsPlus), PNG	115,6	20,38
F6	128 Farben (Dithering, Photoshop), PNG	128,2	22,60
F7	16 Farben (Dithering, Photoshop), PNG	72,3	12,74
F8	24 Bit, (Photoshop) hohe Verluste (Stufe 0), JPEG	36,1	6,36
F9	24 Bit, (Photoshop) geringe Verluste (Stufe 8), JPEG	167,7	29,56
F11	1:2, LWF,	287,5	50,64
F12	1:5, LWF	115,0	20,26
F13	1:10, LWF	57,5	10,13
F14	1:20, LWF	28,7	5,06
F15	1:30, LWF	19,2	3,38
F16	1:40, LWF	14,4	2,54

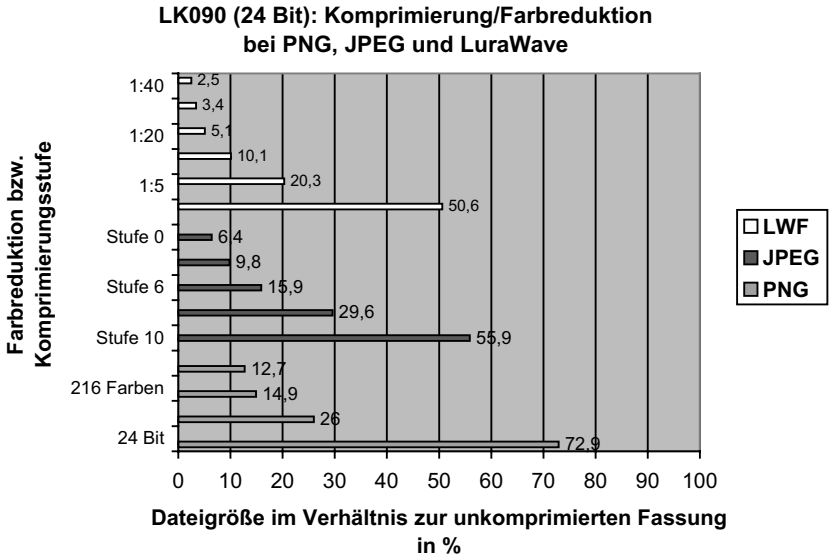


Abb. 8: Testmodell *Majestätssiegel* (Filmdigitalisierung, 24 Bit) – Diagramm Komprimierung/Farbreduktion bei PNG, JPEG und LWF.

des Testmasters gibt es keine merkliche Verschlechterung der Bildqualität in diesem Format.

Die besonderen Eigenschaften des wavelet-basierten LuraWave-Formats werden im Rahmen des Testmodells durch zusätzliche Konversionsformen veranschaulicht (D10). So wurde der direktdigitalisierte TIF-Ausgangsmaster (D0) in eine verlustlos komprimierte LWF-Datei (D10) konvertiert, mit einer Bildgröße von 848 x 878 Pixeln und einer Dateigröße von 1,752 MB. Mit dieser Datei können nun folgende erweiterte Fähigkeiten des LuraWave-Formats demonstriert werden:

- Die Datei muß nicht in ihrer Gesamtgröße geladen werden, sondern es kann die zu ladende Datenmenge vorgegeben werden (hier zunächst 40 KB, d.h. 2,3% der Ausgangsgröße). Der Benutzer kann dann manuell durch Nachladen die Qualität des Bildes verbessern.
- Über eine Zoomfunktion kann die Ansicht on-the-fly auf die Ausgangsgröße der LWF-Datei vergrößert werden. Leider bleibt beim Scrollen des Images die Zoomstufe nicht erhalten.

Bitonale (1 Bit) und Graustufendigitalisierung (8 Bit)

Auch bei der bitonalen (1 Bit) und Graustufendigitalisierung (8 Bit) wurden Bildqualität und Dateigröße in Abhängigkeit von verschiedenen Dateiformaten und Komprimierungen beim PNG-, JPEG- und LWF-Dateiformat untersucht. Zur Visualisierung dieser Untersuchung wurde ebenfalls ein Internet-Testmodell erstellt mit einer Montage der Wochenzeitung *Staatsanzeiger für Baden-Württemberg*.[22]

Für den ersten Test diente als Ausgangsform ein montierter, direktdigitalisierter, unkomprimierter TIFF-Master in 1 Bit Farbtiefe, 300 dpi Scanauflösung, einer Bildgröße von 1375 x 6500 Pixeln und einer Dateigröße von 1,12 MB. Er wurde für die Demonstration auf 330 x 1560 Pixel reduziert und für die Bildschirmdarstellung in eine Farbtiefe von 8 Bit konvertiert. Dabei hat der Testmaster (D0) unkomprimiert eine Dateigröße von 517,6 KB (100%).[23] Die Formatkonvertierung und Bildbearbeitung erfolgte mit ThumbsPlus 4.0 und Adobe Photoshop.

---

22 Siehe Internet-Testmodell *Staatsanzeiger für Baden-Württemberg, 48. Jahrgang, Ausgabe Nr. 1 vom 11.01.1999* unter *http://www.lad-bw.de/digpro/graf.htm*.

23 Im Testmodell wird der Testmaster als PNG-Datei angezeigt, da das TIF-Format in Web-Browsern nicht darstellbar ist.

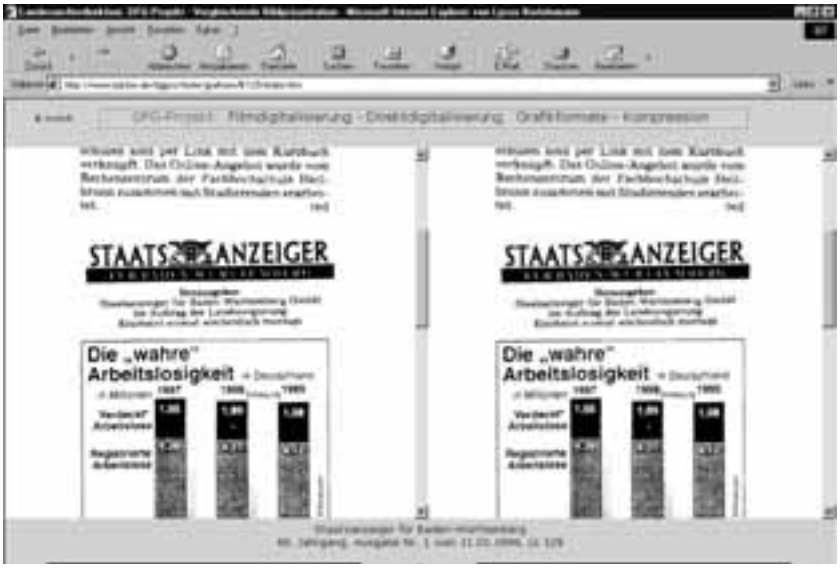


Abb. 9: Testmodell *Staatsanzeiger für Baden-Württemberg (http://www.lad-bw.de/digpro/graf.htm).*

	Testform	KB	%
D0	256 Graustufen, TIF unkomprimiert	517,6	100,00
D1	256 Graustufen, PNG	144,0	27,82
D2	16 Graustufen (Dithering Floyd-Steinberg), PNG	80,1	15,48
D3	8 Graustufen (Dithering Floyd-Steinberg), PNG	64,7	12,50
D4	4 Graustufen (Dithering Floyd-Steinberg), PNG	48,9	9,45
D5	2 Graustufen (Dithering Floyd-Steinberg), PNG	25,0	4,83
D10	Qualität hoch (90%), JPEG	138,1	26,68
D11	Qualität mittel (50%), JPEG	99,5	19,22
D12	Qualität niedrig (5%), JPEG	67,8	13,10
D34	1:2, LWF, embedded	257,4	49,73
D35	1:3, LWF, embedded	171,6	33,15
D36	1:5, LWF, embedded	103,0	19,90
D37	1:10, LWF, embedded	51,5	9,95

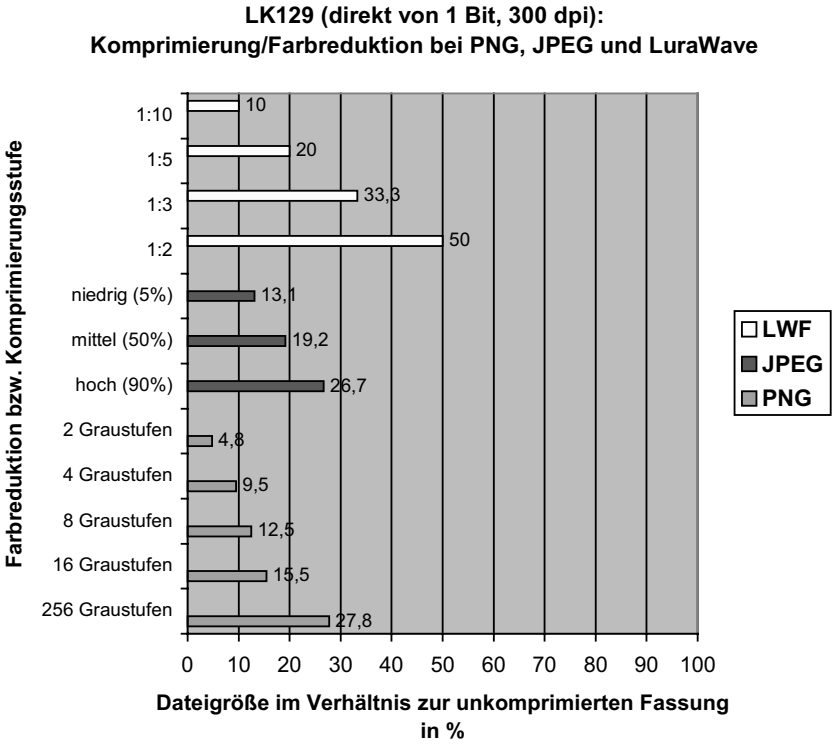


Abb. 10: Testmodell *Staatsanzeiger für Baden-Württemberg* (Direktdigitalisierung mit 1 Bit-Master, 300 dpi) – Komprimierung/Farbreduktion bei PNG, JPEG und LuraWave.

Beim PNG-Format wird mit der Reduktion auf bis zu vier Graustufen (D4, 9,5%) eine noch akzeptable Qualität erreicht. Vergleichbar dazu ist die Form D12 im JPEG-Format. Allerdings ist bei der JPEG-Form eine deutliche Weichzeichnung zu beobachten, die die Lesbarkeit beeinträchtigt. Zudem ist die Komprimierungsrate um 3,6% schlechter als bei der vergleichbaren PNG-Datei. Zieht man zum Vergleich noch das LWF-Format (D37, 10%) hinzu, dann zeigt sich, daß dieses eine dem JPEG-Format vergleichbare Weichzeichnung aufweist, bei einer zur PNG-Form fast identischen Dateigröße.

Der Test macht deutlich, daß für die Aufbereitung von 1 Bit Mastern für eine Internetpräsentation das PNG-Format aufgrund seiner guten Bildqualität (Detailschärfe) am ehesten geeignet ist. Die Formate JPEG und LWF eigenen sich für die bitonale Digitalisierung nur bedingt.

Ein weiterer Test wurde mit einem filmdigitalisierten TIFF-Master (300 dpi, 1375 x 6500 Pixel, unkomprimiert) mit 8 Bit Farbtiefe durchgeführt. Er hat eine Dateigröße von 8,94 MB. Für den Test wurde er auf 330 x 1560

Pixel reduziert. Dabei hat der unkomprimierte Testmaster (F13) eine Dateigröße von 517,6 KB (100%).[24] Die Formatkonvertierung und Bildbearbeitung erfolgte mit ThumbsPlus 4.0 und Adobe Photoshop.

	Testform	KB	%
F13	256 Graustufen, TIF unkomprimiert	517,6	100,00
F14	256 Graustufen, PNG	322,6	62,33
F15	128 Graustufen (Floyd-Steinbg.), PNG	312,6	60,39
F16	64 Graustufen (Floyd-Steinbg.), PNG	282,5	54,58
F17	16 Graustufen (Floyd-Steinbg.), PNG	152,7	29,50
F18	4 Graustufen (Floyd-Steinberg), PNG	80,4	15,53
F23	Qualität hoch (90%), JPEG	172,4	33,31
F24	Qualität mittel (50%), JPEG	120,6	23,30
F25	Qualität niedrig (5%), JPEG	78,7	15,20
F30	1:2, LWF, embedded	257,4	49,73
F31	1:3, LWF, embedded	171,6	33,15
F32	1:5, LWF, embedded	103,0	19,90
F33	1:10, LWF, embedded	51,5	9,95

Beim PNG-Format wird mit der Verkleinerung der Farbtiefe auf bis zu 16 Graustufen (F17) mit einer Reduktion der Dateigröße auf 29,5% eine noch akzeptable Qualität erreicht. Betrachtet man dazu eine vergleichbare Form im JPEG-Format (F24) mit einer Reduktion der Dateigröße auf 23,3%, stellt man fest, daß hier im Gegensatz zu 1 Bit-Ausgangsmastern das JPEG-Format eine um etwa 7% bessere Komprimierung besitzt. Das LWF-Format bringt in der zur Dateigröße vergleichbaren Form (F32) eine Reduktion der Dateigröße auf 19,9%, was im Vergleich zur PNG-Form eine ca. 10% und zur JPEG-Form 3% bessere Komprimierungsrate ergibt.

Betrachtet man die Abbildung 12 (Testmodell *Staatsanzeiger für Baden-Württemberg* (Direktdigitalisierung, 1 Bit und Filmdigitalisierung, 8 Bit) – Komprimierung/Farbreduktion bei PNG, JPEG und LWF) dann zeigt sich, daß bei Verwendung der Dateiformate JPEG und PNG in Graustufen umgewandelte 1 Bit-Scans eine geringere Dateigröße aufweisen als aufbereitete 8 Bit-Scans. 1 Bit-Bilder bieten eine klarere Unterscheidung zwischen Vorder- und Hintergrund, so daß eine höhere Kompression möglich ist. Dabei erreicht man mit dem PNG-Format bei 8 Bit Digitalisaten ähnlich wie bei der Farbdigitalisierung eine größere Konturschärfe als bei vergleichbaren JPEG- und LuraWave-Formen. Die Dateigrößenreduktion wird beim PNG-Format vor allem durch eine Graustufenreduktion erreicht. Sie führt allerdings bei 8 Bit-Scans teilweise zu einem Informationsverlust.

---

24 Ebd.

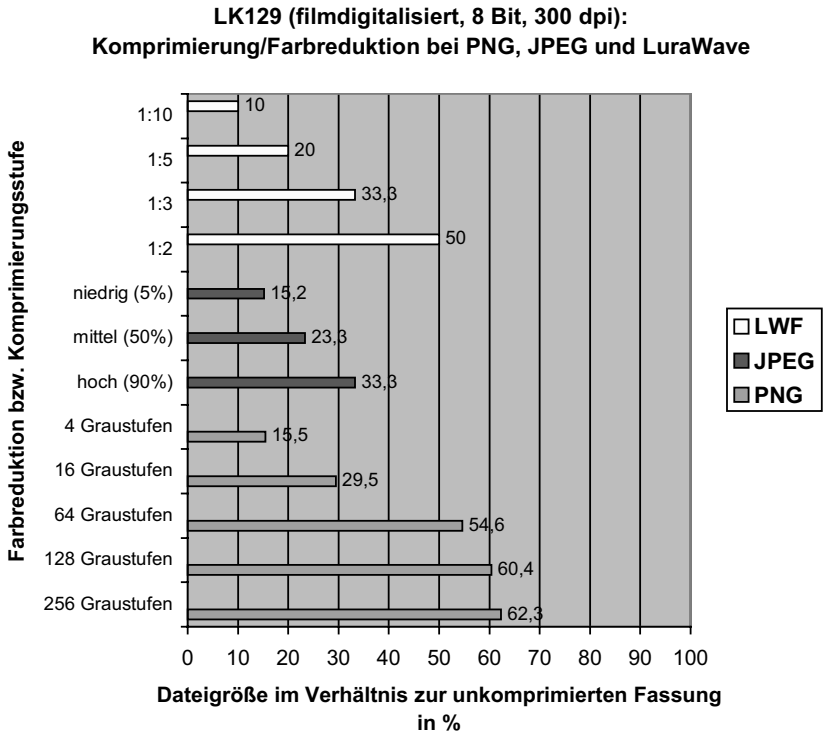


Abb. 11: Testmodell *Staatsanzeiger für Baden-Württemberg (*Filmdigitalisierung mit 8 Bit-Master, 300 dpi) – Komprimierung/Farbreduktion bei PNG, JPEG und *LuraWave.*

Eine Reduktion der Graustufen ist dagegen bei aufbereiteten 1 Bit-Digitalisaten im PNG-Format auf bis zu vier Graustufen (D4) ohne merklichen Informationsverlust möglich. Eine weitere Farbreduktion auf nur zwei Graustufen (D5) und einer weiteren Dateigrößenreduktion führt aber dann zu einer deutlichen Verschlechterung der Bildqualität (Körnungseffekt, Ausbleichen von Schrift), was nicht mehr akzeptabel ist. JPEG- und LuraWave-Formen in vergleichbarer Dateigröße (D12 und D37) reichen qualitativ aufgrund ihrer Artefaktbildung und schlechteren Konturschärfe nicht an die PNG-Form heran.

Zusätzlich zu den im Testmodell dargestellten Formen wurden mit den neuen wavelet-basierten Dateiformaten LuraWave (LWF) und LuraDocument (LDF) weitere Tests durchgeführt, um die unterschiedlichen Eignungen dieser beiden Formen miteinander zu vergleichen. Dafür wurde mit einer weiteren Testmontage des *Staatsanzeigers* eine Testreihe analog zur vorgehenden durchgeführt. Ausgangsform ist ein filmdigitalisierter Testmaster im TIF-Format (300 dpi, 8 Bit Farbtiefe) mit einer Dateigröße von 2,838 MB (100%).

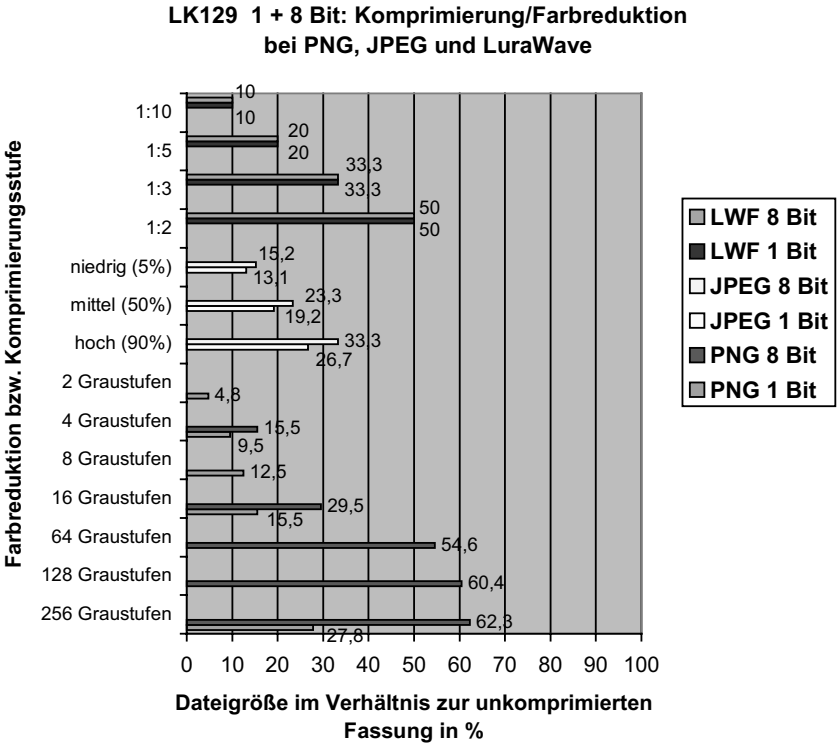


Abb. 12: Testmodell *Staatsanzeiger für Baden-Württemberg* (Direktdigitalisierung, 1 Bit und Filmdigitalisierung, 8 Bit) – Komprimierung/Farbreduktion bei PNG, JPEG und LWF.

Lurawave

Beim LuraWave-Format ist die Textqualität nur bis zu einem Komprimierungsfaktor von 1:20 noch aktzeptabel, die sonstige Bildqualität noch bei 1:40.

Komprimierung	Bildaufbau-Modus	Dateigröße in KB	Dateigröße in %
		2838	100,00
verlustlos	baseline	1977	69,66
verlustlos	embedded	1976	69,63
1:5	baseline	565	19,91
1:5	embedded	568	20,01
1:10	baseline	284	10,01
1:10	embedded	284	10,01
1:20	baseline	142	5,00
1:20	embedded	142	5,00
1:30	baseline	95	3,35
1:30	embedded	95	3,35
1:40	baseline	71	2,50
1:40	embedded	71	2,50
1:40	baseline	57	2,01

Luradocument

Komprimierung Textqualität	Bildqualitätsoption	Dateigröße in KB	Dateigröße in %
		2.838	100,00
1:230	low	37	1,30
1:230	medium	47	1,66
1:230	good	68	2,40
1:230	excellent	101	3,56
1:20	excellent	142	5,00
1:30	excellent	101	3,56
1:920	excellent	101	3,56

Bei Verwendung des LuraDocument-Formats ist der Text noch bei einem Komprimierungsfaktor von 1:920 (Höchstwert) gut lesbar. Der Speicherbedarf sinkt allerdings schon ab 1:30 nicht mehr, die Bildqualität im Umfeld des Textes ist aber selbst bei der Einstellung *excellent* nicht mehr akzeptabel.

Der Vergleich zeigt, daß bei LuraWave mit einer Komprimierung auf 5,0% der Ausgangsgröße (142 KB) eine noch akzeptable Qualität erreicht wird, bei der der Text noch lesbar ist. Bei dem auf Textkomprimierung spezialisierten LuraDocument-Format wird sogar bei einer Reduktion der Dateigröße auf 3,56% (101 KB) der Ausgangsgröße eine noch lesbare Textqualität erzielt. Zum Testzeitpunkt ließ sich LuraDocument aber noch nicht für Internetpräsentationen einsetzen, da ein entsprechendes Browser-Plugin fehlte.

Die Testreihen mit dem Dateiformaten der Firma Luratech verdeutlichen, daß LuraWave für handschriftliches Archivgut besser geeignet ist als LuraDocument. Bei diesen Objekten kann LuraDocument in der Regel Schrift und Bild, wegen der auf Druckschriften optimierten Erkennungsmethode, nicht richtig unterscheiden. Die Übergänge zwischen Schrift und Bild sind technisch gesehen fließend (geringere Kontraste, oft keine monochrome Schrift, helle Schriftfarben, dunkles Papier oder Hintergründe, keine klaren Linien usw.). Bei Vorlagen mit gedrucktem Archiv- und Bibliotheksgut werden mit LuraDocument dagegen bei einer höheren Komprimierung eine bessere Textlesbarkeit erreicht als mit LuraWave. Bei relativ großen und deutlichen Schriften mit LuraDocument kann die Schrift sogar bei sehr hohen Kompressionsraten noch gut gelesen werden.

Kompression und Farbreduktion bei unterschiedlicher Bildbearbeitungssoftware

Es wurde die Erfahrung gemacht, daß mit unterschiedlicher Bildkonvertierungs- und Bildbearbeitungssoftware unterschiedliche Kompressionsraten beim PNG- und JPEG-Dateiformat erzielt werden können. Dies wurde durch eine Testreihe mit den gleichen Masterformen aus den Warenkörben belegt. Folgende Software wurde für den Test eingesetzt:

- ThumbsPlus 3.2 (Cerious Software)
- Photoshop 4.0 (Adobe)
- PhotoImpact 3.0 (Ulead)

Beim PNG-Format unterscheiden sich die Komprimierungsraten von Programm zu Programm. So erreichte z.B. Photoshop 4.0 bei der Konvertierung ins PNG-Format (71,4%, 698 KB) eine um ca. 5% geringe Komprimierungsrate als ThumbsPlus 3.2 (66,2%, 647 KB) bezogen auf die Ausgangsform (TIF, 24 Bit, 478 x 697 Pixel, 977 KB).[25] ThumbsPlus 3.20 und PhotoImpact 3.0 liefern in der Regel ähnlich gute Ergebnisse. Bei Photoshop ist die Komprimierungsrate wesentlich geringer. Es wurde außerdem deutlich, daß die Größe einer Grafikdatei nicht nur von der Farbtiefe im Bild bzw. der Farbreduktion abhängig ist, sondern auch von der verwendeten Ditheringmethode.

PhotoImpact 3.0 kann bitonale Dateien in das PNG-Format umwandeln sowie transparente Bereiche festlegen, Photoshop 4.0/5.0 und ThumbsPlus 3.20 können dies nicht. PixEdit 4.0 unterstützt das PNG-Format überhaupt nicht. PhotoImpact eignet sich also derzeit am besten für die Erstellung von PNG-Dateien.

---

25 Direktdigitalisierte Testform *Ehevertrag Eberhards IV. von Württemberg mit Henriette von Mömpelgard mit Wappen der Häuser Württemberg und Mömpelgard, S. 12* (aus: Hauptstaatsarchiv Stuttgart, A266, U1, Ende 14. Jahrhundert).

Farben	24 Bit		256 Photo		216 WWW Photo		128 Thumbs+		128 Photo		16 Photo	
Formate	KB	%	KB	%	KB	%	KB	%	KB	%	KB	%
unkomprimiert	977	100,0	327	33,5								
TIFF LZW	805	82,4	260	26,6	158	16,2	251	25,7	238	24,4	124	12,7
PNG Thumbs+	647	66,2	218	22,3	141	14,4	197	20,2	159	16,3	109	11,2
PNG PhotoImpact 3.0	664	68,0	218	22,3	140	14,3	211	21,6	197	20,2	107	11,0
PNG Photo 4.0 adaptiv	698	71,4	304	31,1	233	23,8	315	32,2	291	29,8	161	16,5
JPEG Photo Stufe 0	68	7,0										
JPEG Photo Stufe 3	103	10,5										
JPEG Photo Stufe 6	154	15,8										
JPEG Photo Stufe 8	253	25,9										
JPEG Photo Stufe 10	471	48,2										

Photo = Photoshop 5.0; Thumbs+ = ThumbsPlus 3.20
Farbreduktion in Photoshop 5.0 mit Option *Flexibel* und bei *Dithering Diffusion* und *Best*; Farbreduktion in Thumbsplus mit der Option *Dithering Floyd-Steinberg*

Abb. 13: Kompression und Farbreduktion bei unterschiedlicher Bildverarbeitungssoftware.

Zusammenfassung

Eine verbindliche Empfehlung für eines der genannten Dateiformate als Benutzungsversion kann zum jetzigen Zeitpunkt nicht gegeben werden, da die jeweilige Auswahl von Fall zu Fall durch die Art der Vorlagen (Text/Strich, Halbton, Farbe) mitbestimmt wird.

Für digitale Farb- und Graustufen-Master (8–36 Bit) bietet sich weiterhin das TIF-Format in seiner unkomprimierten Variante als momentaner Quasi-Standard an, während es als TIFF-G4 mit verlustfreier Kompression für 1 Bit-Master sinnvoll erscheint. Die Lage könnte sich aber mit einer weiteren Verbreitung des PNG-Dateiformats ändern, das ebenfalls aufgrund seiner verlustfreien Kompression und anderer Vorteile nicht nur ein zukünftiges Dateiformat für digitale Nutzungsformen werden könnte, sondern bei entsprechender Weiterentwicklung (z.B. Möglichkeit der Einbindung von Farbprofilen) auch für digitale Master. Im Hinblick für die Langzeitarchivierung digitaler Farbmaster könnte auch das von der Fa. Kodak entwickelte ImagePac-

Format mit Unterstützung des ebenfalls von Kodak propagierten geräteunabhängigen YCC-Farbraums eine Alternative bieten, das im Rahmen dieses Projektes nicht getestet wurde.

Bei der Erstellung von digitalen Nutzungsformen für Präsentationen ist man immer noch mit Verwendung des JPEG-Dateiformats auf der sichersten Seite, da dieses Format auch von älterer Browser-Software gelesen werden kann. Das GIF-Format (Graphics Interchange Format) – geeignet für 1 Bit-Bilder und für 8 Bit-Graustufen- und Farbimages – wird aufgrund seines lizenzpflichtigen Komprimierungsalgorithmus nicht empfohlen. Die vorteilhafte Eigenschaft des GIF-Formats gegenüber dem JPEG-Format, nämlich die Bewahrung einer guten Konturschärfe bei Schriftvorlagen trotz Kompression, besitzt auch das universalere PNG-Dateiformat. Farbreduzierung und PNG-Format sind vor allem für die Aufbereitung von 1 Bit-Digitalisaten gegenüber dem JPEG-Format vorzuziehen, da hier durch Graustufenreduktion mit dem PNG-Format deutlich geringere Dateigrößen erzielt werden können als bei 8bit-Digitalisaten.

Daneben gibt es die bisher noch proprietären Dateiformate mit Wavelet-Komprimierungsalgorithmen, die bei akzeptabler Qualität eine beachtliche Komprimierung erreichen. Diese Formate sind besonders dann geeignet, wenn Bilder in verschiedenen Qualitätsstufen angezeigt werden sollen, ohne daß dafür mehrere einzelne Dateien in unterschiedlicher Qualität und Dateigröße aufbereitet werden müssen. Die neuen wavelet-basierten Dateiformate, von denen die Formate LuraWave und LuraDocument getestet wurden, werden sich wohl erst dann stärker verbreiten, wenn sich ein gemeinsamer Standard, d.h. ein Dateiformat, am Markt durchgesetzt hat oder eine Integration der Dateiformate in die gängigen Web-Browser stattgefunden hat, ohne daß gesonderte Plug-ins installiert werden müssen. Bisher gibt es hier nur proprietäre Lösungen, so daß noch keine Empfehlung in diesem Bereich ausgesprochen werden kann.

5. *Image Enhancement* und Bildaufbereitung

Nach der Qualitätskontrolle sind die nächsten Schritte innerhalb eines Digitalisierungs-Workflows das sog. Image Enhancement und die Aufbereitung des Bildmaterials für Präsentationen z.B. im Internet.

Bereiche und besondere Techniken des Image Enhancements

Wichtige Bereiche des *Image Enhancements* sind:

- Tonwertkorrektur:
  - Höhen und Tiefen
  - Kontrast
  - Gamma

- Farbkorrektur:
  - Sättigung
  - Farbton
  - Farbstich entfernen: mit *Farbbalance* oder mit *farbkanalabhängigen Gradationskurven*
- Schärfen-Korrektur:
  - durch Filter *Unscharf maskieren*
  - oft läßt sich die Schärfe auch durch *Herunterrechnen* der Auflösung und anschließendes Scharfzeichnen verbessern
- Beseitigung von Bildstörungen (z.B. Entfernen von Moiré)
- Korrektur der Bildlage von schiefen Bildern
- Entfernen unerwünschter Ränder
- Trennung doppelseitiger Bilder
- Verkleinern von Bildern

Trennung von doppelseitigen Bildern

Die Trennung von doppelseitigen Bildern ist für Bücher und mehrseitige Akten von Bedeutung. Dabei hat eine Seitentrennung folgende Vorteile:

- bessere Präsentationsmöglichkeit mit Webbrowser-Technik: nur vertikales Scrollen und kein horizontales Scrollen, um eine Lesbarkeit von Text zu gewährleisten,
- einfachere Bildverwaltung, da jede Seite über eine eindeutige Kennzeichnung oder über die Seitenzahl angesteuert werden kann.

Grundsätzlich gibt es zwei Möglichkeiten für eine Seitentrennung:

1. Die Seitentrennung erfolgt durch Dienstleister im Rahmen des Scanvorgangs mit entsprechender Scansoftware oder
2. im Rahmen einer Nachbearbeitung über Makros und Batch-Funktionen von Bildbearbeitungssoftware. Allerdings ist ein vollautomatisches Verfahren schwierig mit Standardsoftware zu realisieren. Für die Benennung der Dateinamen sind manuelle Eingriffe in den Stapelablauf notwendig.

Praktisch erprobt wurde die Seitentrennung am Internetpräsentationsmodell des Buches *Rudolf Krauß, Das Stuttgarter Hoftheater von den ältesten Zeiten bis zur Gegenwart, Stuttgart 1908.*[26]

---

[26] URL: *http://www.lad-bw.de/digpro/praes.htm.* Vgl. dazu in diesem Band den Beitrag von August *Wegmann.*

Entfernung von schwarzen Bildrändern

Falls größere schwarze Ränder vorhanden sind, können diese manuell oder, wenn die Seiten innerhalb der Bildfläche an einem Fixpunkt ausgerichtet sind, mit einer automatischen Stapelbearbeitung über ein Bildverarbeitungsprogramm wie Photoshop mit *Freistellen*-Werkzeug und entsprechenden Makros entfernt werden. Schwarze Ränder führen zu einer unnötigen Erhöhung der Dateigröße und sind auch störend für Bildschirmpräsentationen.

Korrektur der Bildlage von schief eingescannten Vorlagen

Eine Korrektur der Bildlage von schief eingescannten Vorlagen wurde mit Adobe Photoshop 4.0/5.0 über Funktion *Transformieren/drehen* unter Zuhilfenahme von eingeblendeten Hilfslinien (Linialfunktion) realisiert. Dabei ist zu beachten, daß die Transformierung in der Originalgröße vorgenommen wird und die Bilder erst im Anschluß auf die benötigte Bildgröße verkleinert werden.

Entfernung von Streupixeln in bitonalen Scans

Eine Verbesserung des visuellen Eindrucks von 1 Bit-Scans wird durch die Entfernung von störenden Streupixeln erreicht. Sie kann über spezielle Bildbearbeitungssoftware wie PixEdit 4.0 erfolgen. Dabei bietet das Programm eine Makrofunktionalität mit der Möglichkeit der Abarbeitung der Makros im Batch-Betrieb.

Tests wurden an Objekten des *Kleinen Warenkorbs* und an ausgewählten Stücken des *Großen Warenkorbs,* wie dem Buchmodell *Stuttgarter Hoftheater,*[27] vorgenommen.

Die verschiedenen Arbeitschritte der Aufbereitung digitaler Masterformen für Präsentationen

Die Aufbereitung digitaler Masterformen für Präsentationen besteht in erster Linie in einer Anpassung der Bildgröße für die Bildschirmdarstellung mit einer gleichzeitigen Reduktion der Dateigröße und Konvertierung in ein für Webbrowser darstellbares Dateiformat.

---

[27] Ebd.

Ein solcher Workflow sieht mit der Bildverarbeitungssoftware Adobe Photoshop vereinfacht folgendermaßen aus:

1. Ändern der Bildgröße (ca. 500 Pixel Breite für eine Bildschirmauflösung von 800 x 600 Pixel)
2. Eventuell Bild freistellen mit Schneidewerkzeug
3. Eventuell Tonwertkorrektur über Histogramm-Funktion oder Gamma-Kurve
4. Bild schärfen mit dem Filter *Unscharf maskieren*
5. Eventuell Korrektur der Farbsättigung
6. Bild speichern im PNG- oder JPEG-Format

Im folgenden werden Beispiele und Besonderheiten dieses Workflows gesondert für Farb-, Graustufen- und bitonale Masterformen vorgestellt.

Aufbereitung von Farbbildern (24 Bit)

Hauptbereiche der Farbbildbearbeitung sind Tonwertkorrektur, Korrektur der Farbwiedergabe und Korrektur der Bildschärfe.

1. Die Veränderung des Tonumfangs bzw. Korrektur der Helligkeitsverteilung bei digitalen Farbbildern erfolgt entweder mittels eines Histogramms (Verteilung der Helligkeitsstufen 0–255 eines Bildes), einer Tonwertkurve oder über *Pipetten*. Dabei könnten folgende Einstellungen verändert werden:

   a) Weißpunkt (hellster Punkt im Bild),
   b) Schwarzpunkt (dunkelster Punkt im Bild),
   c) Gammawert[28] und
   d) Kontrast.

   Das Ziel der Tonwertkorrektur ist die gleichmäßige Verteilung der Töne über den gesamten verfügbaren Tonumfang, so daß möglichst viele Tonabstufungen der Vorlage reproduziert werden.
2. Die Korrektur der Farbwiedergabe erfolgt durch Änderung der Farbbalance mit dem Ziel, eine ausgewogene, der Vorlage entsprechende Farbwiedergabe zu erreichen.
3. Die Korrektur der Bildschärfe erfolgt durch Einsatz von Korrekturfiltern wie *Unscharf maskieren*.

---

[28] Eine Veränderung des Gammawerts führt zu einer Veränderung der Bildhelligkeit durch eine Neuverteilung der Helligkeitswerte (Einfluß auf Mitteltöne), ohne daß der Schwarz- und Weißpunkt verschoben wird.

Am Internet-Testmodell *Majestätssiegel*[29] können die verschiedenen Arbeitsschritte der Farbbildbearbeitung demonstriert werden. Der filmdigitalisierte Master wurde hierbei für die Darstellung im Webbrowser aufbereitet:

F0: 24 Bit-TIFF-Master, unbearbeitet[30]
F1: Bearbeitete Fassung im PNG-Format mit 24 Bit-Farbtiefe nach der Verkleinerung des Masters und durchgeführter Schärfenkorrektur
F2: Bearbeitete Fassung nach durchgeführter Farb- und Tonwertkorrektur
F3: Bearbeitete Fassung nach durchgeführter Farbreduktion auf 256 Farben

Optimal bei der Bildaufbereitung ist die Kombination verschiedener Bildverarbeitungssoftware. In diesem Fall wurden die beiden Programme Adobe Photoshop 5.0 und ThumbsPlus 3.30 miteinander kombiniert. So bringt z.B. die Farbreduktion in Photoshop auf 16 Farben bessere Ergebnisse als die mit ThumbsPlus. Umgekehrt sind die Resultate bei der Konvertierung ins PNG-Format mit ThumbsPlus besser als mit Photoshop. Mit ThumbsPlus wird eine höhere Kompression und bessere Darstellungsqualität in Webbrowsern erreicht als mit Photoshop 5.0 (dunkle Darstellung mit stark erhöhter Sättigung). Letztere Software kann das PNG-Format bisher nicht richtig verarbeiten. Eine Reduktion von 16,7 Mio (24 Bit) auf 256 bzw. 64 Farben ist für viele Farbvorlagen akzeptabel, dagegen ist eine solche auf 16 Farben (4 Bit) nicht mehr akzeptabel, da die unvergleichbar höhere Reduktion durch Farbausfall zu einer Veränderung von Farbe und Helligkeit führt.

Aufbereitung von Graustufenbildern (8 Bit)

Bei 8 Bit Scans sind folgende Schritte der Bildaufbereitung für die Online-Präsentation mit Photoshop 5.0 notwendig:

1. Verkleinern
2. Scharfzeichnen mit Filter *Unscharf maskieren*
3. Eventuell Tonwertkorrektur
4. Graustufenreduktion

---

29 Siehe Internet-Testmodell *Majestätssiegel* unter *http://www.lad-bw.de/digpro/graf.htm*.
30 Im Testmodell wird der Testmaster als PNG-Datei angezeigt, da das TIF-Format in Web-Browsern nicht darstellbar ist.

Eine Reduktion der Graustufen bei 8 Bit auf 4 Bit wirkt sich visuell nicht so deutlich aus wie eine Farbreduktion bei Farbbildern. Ein 24 Bit-Farbbild enthält bekanntlich dreimal so viel Information wie ein 8 Bit-Graustufenbild.

Am Internetpräsentationsmodell für das Online-Buch *Christoph Weigel, Ein Schock Phantast'n in einem Kasten mit Ihrem Portrait gar net in Kupffer gebracht und ausgelacht. – Nürnberg : Weigel [circa 1690]*[31] kann die Aufbereitung von Graustufenbildern demonstriert werden.

Insgesamt handelt es sich um 75 Buch-Doppelseiten, die als filmdigitalisierte TIFF-Master, 8 Bit, Bildgröße ca. 4000 x 2400 Pixel, 300 dpi Scanauflösung bezogen auf die Vorlage, mit jeweils ca. 9,5 MB Dateigröße vorliegen.

Eine Bildserie im PNG-Format wurde aus den Masterformen mit dem Programm ThumbsPlus 3.30 im Batch-Betrieb folgendermaßen erstellt:

1. Reduktion der Bildgröße auf 50% mit Einstellung *best fit for Sizing*
2. Scharfzeichnen mit Filter *sharpen less*
3. Reduktion der Farbtiefe auf 16 Graustufen (4 Bit)
4. mehrere Durchgänge *sharpen less*
5. Abspeichern im PNG-Format

Durch diese Bearbeitungsschritte entstanden digitale Nutzungsformen, die für Webbrowser geeignet sind, mit einer Dateigröße von jeweils 570 KB.

Aufbereitung von bitonalen Bildern (1 Bit)

Bitonal gescannte Bilder müssen für die Präsentation für die Darstellung in Webbrowsern und die Bearbeitung mit Filtern in 8 Bit-Graustufenbilder umgewandelt werden.[32]

Die Aufbereitung von 1 Bit-Mastern kann am Beispiel des Internet-Buchmodells *Stuttgarter Hoftheater*[33] demonstriert werden. Die insgesamt 361 1 Bit-TIFF-Master mit einer Dateigröße von jeweils etwa 474 KB wurden für eine Präsentation mit Photoshop 5.0 und ThumbsPlus im Batch-Betrieb aufbereitet:

---

31 URL: *http://www.lad-bw.de/digpro/praes.htm*. Vgl. dazu in diesem Band den Beitrag von August *Wegmann.*

32 Dasselbe gilt für Bilder mit *indizierten Farben* (z.B. 4 Bit mit 16 Graustufen). Sie müssen bevor sie bearbeitet werden können, zuerst wieder in den 8 Bit-Modus umgewandelt werden.

33 Siehe dazu *http://www.lad-bw.de/digpro/praes.htm* und den Beitrag in diesem Band von August *Wegmann*.

1. Entfernen von Störpixeln mit Filterfunktion der Software PixEdit in den 1 Bit-Masterformen
2, Umwandlung der Masterformen in 8 Bit-Graustufenbilder (Zwischen-Master) im Größenverhältnis 1:1
3. Größenreduktion des Zwischenmasters auf eine bildschirmlesbare Größe
4. Freistellen, d.h. entfernen des unnötigen Randes und Korrektur der Bildlage
5. Scharfzeichnung mit Filter *Unscharf maskieren* nach visuellem Eindruck
6. Abspeichern in verschiedenen Dateiformaten:
   – PNG mit Reduktion der Farbtiefe auf 16 Bit (64 Graustufen): 179 KB Dateigröße je Image
   – JPEG mit Komprimierungsstufe: 275 KB Dateigröße je Image

Ergebnisse bei der Bearbeitung von 1 Bit und 8 Bit Digitalisaten

Bei der Erstellung der Internet-Testmodelle für den Vergleich verschiedener Dateiformate und Komprimierungsverfahren können zusätzlich zu den dokumentierten Testergebnissen zum Thema *Bildkomprimierung und Darstellungsqualität* folgende Ergebnisse für die Bildaufbereitung festgehalten werden:

- Bitonale Scans (1 Bit) benötigen weniger Nachbearbeitung als Graustufenscans (8 Bit). Wichtig ist allerdings, daß der Schwellenwert richtig gewählt wurde. Wurde er falsch gesetzt, ist durch Nachbearbeitung keine Verbesserung mehr zu erreichen.
- Bitonale Scans können, wenn sie im Browser dargestellt werden sollen, nicht im bitonalen Modus belassen werden. Sie müssen vor der Größenanpassung in den Graustufenmodus konvertiert werden. Die Graustufen können anschließend jedoch bis auf vier Stufen reduziert werden, ohne daß größere Qualitätseinbußen bei der Schrift hinzunehmen sind.
- Mit 1 Bit Digitalisaten von 600 dpi-Master-Formen werden trotz Verkleinern der Images und Reduktion auf die Bildschirmauflösung eine bessere Bildqualität erzielt als mit solchen von 300 dpi-Mastern.[34]
- Bei Graustufenscans (8 Bit) kann durch Nachbereitung eine größere Qualitätsverbesserung erreicht werden als bei 1 Bit-Scans.

---

[34] Siehe vergleichendes Testmodell *Staatsanzeiger für Baden-Württemberg* unter *http://www.lad-bw.de/digpro/graf.htm.*

- Die Anzahl der Graustufen kann ohne merkliche Qualitätseinbußen bis auf 16 reduziert werden. Bei 4 Graustufen kann es zu deutlichen Bildverschlechterungen durch Informationsverlust kommen.

Automatisierte Bildaufbereitung für Internetpräsentationen von sequentiellen Buchobjekten

Ein wichtiger Bereich des *Image Enhancements* und der Bildaufbereitung ist die Automatisierung dieser Vorgänge. Sie ist aus Gründen der Wirtschaftlichkeit geboten. So kann eine automatische Bildaufbereitung für die Internetpräsentation von sequentiellen Objekten wie Büchern mit doppelseitiger Aufnahme folgendermaßen ablaufen:

1. Duplizieren der Bilder für die Bearbeitung
2. Teilen der Doppelseiten (dadurch entstehen aus jeder Bilddatei zwei neue)
3. Eventuell Geraderichten der einzelnen Seiten. Hier ist ein manueller Eingriff nötig, da jede Seite visuell über die Bildbearbeitungssoftware geradegerichtet werden muß.
4. Reduktion der Bildgröße
5. Tonwertkorrektur
6. Scharfzeichnen
7. Eventuell Reduktion der Farbtiefe
8. Speicherung der Dateien in einem Internet tauglichen Dateiformat unter Vergabe von Dateinamen mit Numerierung zur eindeutigen Identifikation jedes Images und zur Verknüpfung mit Erschließungsdaten

Die Umsetzung dieses Workflows wurde mit marktgängiger Standardsoftware erprobt, unter Nutzung der programmeigenen Makro- und Batchfunktionen. Getestet wurden dabei hauptsächlich die folgenden Programme:

Programm	Schwerpunkte und Funktionen für automatisierte Bildaufbereitung	Fehlende Funktionen für automatisierte Bildaufbereitung
Adobe Photoshop 4.0/5.0	– Bildbearbeitung für Druckvorstufe – Makrorecorder (*Aktionen*) und Batchfunktionalität	– keine automatische Modifizierung von Dateinamen – keine internetgerechte Dateigrößenreduktion
Adobe Image Ready 1.0	– Bildbearbeitung für Internet – Makrorecorder (*Aktionen*) und Batchfunktionalität Arbeitsfläche nachProzent-	– Größe des Auswahlrechtecks kann nicht vordefiniert werden – keine Möglichkeit die angaben zu verkleinern – keine automatische Modifizierung von Dateinamen – Performance-Probleme
Cerious ThumbsPlus 3.2/4.0	– Bildverwaltung mit Bildbearbeitungsfunktionen – Batchfunktionalität ecks	– Größe des Auswahlrechtecks kann nicht vordefiniert werden – keine Möglichkeit der freien Bildrotation mit Rahmen – keine Möglichkeit zur Bildduplizierung
Equilibrium DeBabelizer 4.51	– Dateikonvertierung – umfangreiche Scriptsprache und Batchfunktionalität	– keine Filter für Scharfzeichnen – Größe des Auswahlrechtecks kann nicht vordefiniert werden – keine Möglichkeit der freien Bildrotation mit Rahmen – keine Möglichkeit zur Bildduplizierung
Ulead Photo Impact 4.2	– Bildbearbeitung mit Spezialwerkzeugen für Internet (*Smartsaver*)	– geringe Batchfunktionalität

Keines dieser Programme verfügt über alle für die automatisierte Bildaufbereitung notwendigen Funktionen, so daß sie miteinander kombiniert werden müssen und die Bildaufbereitung in verschiedenen Etappen erfolgen muß.

Zunächst können in *Photoshop* unter Verwendung der Makrofunktionen (*Aktionen*) folgende Arbeitsschritte ausgeführt werden:

1. Duplizieren der Bilder für die Bearbeitung
2. Teilen der Doppelseiten. Alle rechten und alle linken Buchseiten werden in einem gesonderten Ordner abgelegt
3. Eventuell Geraderichten der einzelnen Seiten. Hier ist ein manueller Eingriff in den Stapelablauf nötig, da jede Seite visuell über die Bildbearbeitungssoftware geradegerichtet werden muß.
4. Reduktion der Bildgröße
5. Tonwertkorrektur
6. Scharfzeichnen (Unscharf maskieren)
7. Eventuell Reduktion der Farbtiefe

Mit der Freeware-Software *Rename-It* werden die Dateinamen geändert, so daß linke und rechte Buchseiten zu unterscheiden sind und alle Bilder zur Weiterverarbeitung wieder zusammen in einen Ordner abgelegt werden können.

Mit *ThumbsPlus* oder *DeBabelizer Pro* werden die Bilder dann in einem Internet tauglichen Dateiformat wie PNG oder JPEG im Batchbetrieb abgespeichert. Bei Verwendung des PNG-Formats wird zuvor eine Farbtiefenreduktion durchgeführt, bei Verwendung des JPEG-Formats wird eine geeignete Qualitätsstufe ausgewählt.

6. Die objekt- und mediengerechte Präsentation von Archiv- und Bibliotheksgut

Für die objekt- und mediengerechte Präsentation der aufbereiteten Bilder wurden verschiedene Modelle geschaffen mit dem Ziel, die erzeugten Bilddaten mit archivischen und bibliothekarischen Metadaten wie Titelaufnahme, strukturellen Erschließungsdaten bzw. Dokumentenstrukturdaten (Imagezählung, Seitenzählung innerhalb sequentieller Images etc.) und technischen Metadaten (z.B. Verknüpfungsdaten, Datei-Header-Daten) zu verknüpfen.

Konzeption und Formen von Präsentationsmodellen

In Kooperation mit der Archivschule Marburg wurde ein Fachkonzept zur formalen und inhaltlichen Aufbereitung der unterschiedlichen Gattungen von Archiv- und Bibliotheksgut für den Zugriff im Internet erarbeitet.[35]

---

[35] Siehe dazu in diesem Band den Beitrag von Karsten *Uhde:* Kontextbezogene Online-Präsentationen von Archivgut. S. 233.

Gegenüber einer konventionellen Präsentation und Nutzung von Archivalien kann bei Online-Präsentationen ein sogenannter Mehrwert realisiert werden. Er ergibt sich unter anderem durch

- erweiterte Zugänglichkeit unabhängig von Ort und Zeit,
- einen mehrdimensionalen Zugriff,
- erweiterte Navigationsmöglichkeiten,
- Erschließung auf Dokumentenebene im Kontext,
- Integration der Dokumente in Tektonik und Kontext (Kommentare, didaktische Aufbereitung) und
- Lesehilfen z.B. für eine zeilengetreue Transkription.

Auf der Grundlage des Fachkonzepts wurden verschiedene Formen von Präsentationsmodellen unter Verwendung von Internet-Technologie (HTML, DHTML, Java Script) entwickelt. Dabei wurden Darstellungsqualität, technische Realisierung und Wirtschaftlichkeit berücksichtigt.

Insgesamt kann dabei zwischen zwei Gruppen von Modellen unterschieden werden:

1. individuell gestaltete Präsentationsmodelle ausgewählter Archivalien mit besonderer inhaltlicher Erschließung für Unterrichtszwecke. Berücksichtigt werden dabei erstens Benutzerbedürfnisse, zweitens archivfachliche Gesichtspunkte und drittens die mit der Internet-Technologie verbundenen neuen Möglichkeiten eines mehrdimensionalen Zugriffs. So beinhalten die Präsentationsmodelle u.a. eine automatische Transkriptionshilfe für handschriftliche Archivalien, verschiedene Vergrößerungsstufen, Indexfunktionen, Zoomfunktionen und Hilfefunktionen.
2. standardisierte Modelle für die Präsentation der verschiedenen Gattungen von Archiv- und Bibliotheksgut als unterste Stufe einer kontextorientierten Erschließung.

Außerdem wurden auch die Möglichkeiten einer rationellen Einbindung der digitalen Konversionsformen mittels Makro- und Batchverarbeitung erprobt wie

- automatische Formatänderung (TIFF nach JPEG/PNG),
- automatische Größenanpassung,
- automatische Tonwertkorrektur oder
- automatisches Bildschärfen.

Individuell gestaltete Präsentationsmodelle

Für die zweite Gruppe von Präsentationsmodellen wurden digitalisierte Archivalien in unterschiedlicher Form (Akten, Urkunde, Karten und Pergament-Rotulus) ausgewählt, und objektgerecht mit didaktischer Zielsetzung für einen mehrdimensionalen Zugriff aufbereitet.[36]

Modell 1: Prozeß gegen Ludwig Pfau (Redakteur des politischen Satiremagazins *Der Eulenspiegel*), 1848
Modell 2: Königsurkunde für das Kloster Gottesaue, 1110
Modell 3: Der Fall des Mannheimer jüdischen Rechtsanwalts Waldeck, 1933
Modell 4: Augenscheinkarte mit Landschaft des Taubergrunds bei Laudenbach und Rinderfeld im heutigen Main-Tauber-Kreis, 1575
Modell 5: Große kurpfälzische Rheinstromkarte, 1595
Modell 6: Habsburger Einkünfterodel, um 1291

Texte und Überschriften wurden bei den Präsentationen bewußt kurz gehalten. Die Fließtexte wurden in klare kleine Einheiten aufgeteilt. Kernstück einer jeden Präsentation sind die Quellenseiten. Dabei werden die Quellen immer in mehreren Ansichten (Gesamtansicht, lesbare Detailansichten, Transkriptionen, Übersetzungen) angeboten. In der Gesamtansicht der Quelle gibt es oft die Möglichkeit durch einen Klick mit dem Mauszeiger zu einer vergrößerten Ansicht oder einem vergrößerten Ausschnitt zu gelangen.

Standardisierte Modelle für die Präsentation der verschiedenen Gattungen von Archiv- und Bibliotheksgut

Bei der zweiten Gruppe von Präsentationsmodellen, den standardisierten Modellen, werden neben der Realisierung verschiedener Navigations- und Recherchefunktionen v.a. die Möglichkeiten einer Einbindung der digitalen Konversionsformen mittels Makro- und Batchverarbeitung erprobt.

Ein wichtiges Ziel ist bei diesen Modellen die rationelle Verbindung von digitalisierten Formen mit archivischen Findmitteln und ihre Einbindung als unterste Stufe eines archivischen Online-Angebots.

Bisher wurden folgende Prototypen entwickelt:

---

36 URL: *http://www.lad-bw.de/digpro/praes.htm*. Die Modelle wurden von Dr. Thomas Fricke auf Werkvertragsbasis in Zusammenarbeit mit dem Projektteam erstellt. Siehe dazu in diesem Band den Beitrag von Thomas *Fricke:* Didaktische Internet-Präsentationen von Archivgut. S. 247.

- eine Beispielsammlung für die Präsentation verschiedener Gattungen von Archivgut wie Amtsbücher, Akten, Urkunden, Siegel, Karten, Wappen und Photographien,[37]
- ein Präsentationsmodell für mehrseitige Objekte wie Bücher oder Akten[38] und
- ein Modell für die Präsentation großformatiger Zeitungen.[39]

Alle Prototypen werden in einem Folgeprojekt für den praktischen Einsatz und für die Mengenverarbeitung weiterentwickelt.

Ausblick

Mit dem Projekt wurden zahlreiche Möglichkeiten aufgezeigt, mit Standardwerkzeugen der Bildbearbeitung sowohl die vom Mikrofilm und zum Vergleich vom Original digitalisierten Rohdaten zu hochwertigen digitalen Mastern zu optimieren, als auch zu digitalen Nutzungsformen für Internetpräsentationen aufzubereiten. Versuche mit verschiedenen Dateiformaten, Farbtiefen und Scanauflösungen haben gezeigt, daß hier objektbezogen unterschiedliche Auswahlen getroffen werden müssen, um eine optimale Bildqualität und rationelle Verarbeitungsverfahren zu erreichen.

Da die digitale Bereitstellung von Archiv- und Bibliotheksgut zugleich ein Massenproblem darstellt, gilt es die Arbeitsabläufe zu rationalisieren, um sie möglichst kostengünstig zu gestalten. Für die Bildbearbeitung und digitale Bereitstellung müssen daher einerseits Qualitäts- und Bearbeitungsrichtlinien und andererseits nachnutzbare Werkzeuge in Form von Makros oder Softwareanpassungen entwickelt und in optimierte und standardisierte Arbeitsabläufe integriert werden. Erste Ansätze in dieser Richtung konnten in dem Projekt mit der Erarbeitung von Qualitätskriterien und der Entwicklung von Makros für automatisierte Verarbeitungsverfahren realisiert werden.

Insgesamt eröffnet die retrospektive Digitalisierung – gerade in Form der Filmdigitalisierung – in Verbindung mit der Bereitstellung digitaler Findmittel im Internet bisher nicht dagewesene Möglichkeiten und Qualitäten für die Nutzung von Archiv- und Bibliotheksgut.

---

37 URL: *http://www.lad-bw.de/digpro/praes.htm.* Das Modell wurde in Kooperation mit Herrn Dr. Karsten Uhde und Vinzenz Lübben von der Archivschule Marburg entwickelt. Siehe dazu in diesem Band den Beitrag von Karsten *Uhde.*

38 URL: *http://www.lad-bw.de/digpro/praes.htm.* Das Modell wurde in Kooperation mit Prof. August Wegmann, Fachhochschule Karlsruhe, Fachbereich Wirtschaftsinformatik entwickelt. Siehe dazu in diesem Band den Beitrag von August *Wegmann.*

39 URL: *http://www.lad-bw.de/digpro/praes.htm.* Das Modell wurde durch das Projektteam konzipiert und programmtechnisch von Dr. Thomas Fricke umgesetzt. Siehe dazu in diesem Band den Beitrag von Thomas *Fricke:* Internet-Präsentationsmodelle für Zeitungen und Großformate. S. 297.

Colormanagement bei der Farbdigitalisierung von Archivgut
Grundlagen, Hintergründe und Ausblick

Von Gerald Maier

Einführung

Bei der Farbdigitalisierung von Archivgut stellt die authentische Wiedergabe von Farben eine besondere Herausforderung dar.[1] Dabei spielt es keine Rolle, ob ein Objekt direkt oder über den Zwischenschritt einer Mikroverfilmung vom Film digitalisiert wird. Eine der Originalvorlage entsprechende, exakte Farbwiedergabe ist immer dann erforderlich, wenn dies der *intrinsische Wert*, aber auch der *Evidenzwert* und *Informationswert* des zu digitalisierenden Objektes verlangen.

Besitzt Archivgut einen *intrinsischen Wert,*[2] d.h. einen eigenständigen, formalen Wert mit eigener Aussagekraft, vermittelt es Aussagen auf nicht-textlichem Wege. So bezeugen z.B. bestimmte Farben einer Pergamenturkunde, eines Amtsbucheinbandes oder von Geschäftsgangvermerken auf Aktenschriftstücken die Authentizität oder Herkunft der Objekte. Die Farbdigitalisierung mit möglichst exakter Farbwiedergabe kann mit dazu beitragen, daß dem intrinsischen Wert bei der Benutzung von Archivgut Rechnung getragen wird. Ungeachtet dessen können natürlich auch digitale Konversionsformen in Farbe von Archivalien mit intrinsischem Wert nicht das Originalobjekt ersetzen, sondern nur die Qualität der Nutzungsformen erhöhen.

---

1 Der folgende Beitrag ersetzt den beim Kolloquium gehaltenen Vortrag von Lars Raffelt, Doerner Institut der Bayerischen Staatsgemäldesammlungen München, mit dem Thema *Digitale Photographie und Reproduktion: Konzepte und Fehlerquellen.* Dabei gehen die Ausführungen zum Thema *Colormanagement* und dessen Bedeutung für die Farbdigitalisierung von Archivgut über den Vortrag von Lars Reffelt hinaus, dem es aus zeitlichen Gründen nicht möglich war, zu seinem Vortragsthema einen Beitrag zu verfassen.

2 Siehe dazu Angelika *Menne-Haritz* und Nils *Brübach:* Der intrinsische Wert von Archiv- und Bibliotheksgut. Kriterienkatalog zur bildlichen und textlichen Konversion bei der Bestandserhaltung (Veröffentlichungen der Archivschule Marburg 26). Marburg 1997. Besonders S. 22ff. und *http://staff-www.uni-marburg.de/~mennehar/germanterms.htm*, Art. Intrinsischer Wert. Der *intrinsische Wert* liefert Kriterien für die Anwendung der verschiedenen Verfahren der Bestandserhaltung von der Originalerhaltung bis zur Ersatzverfilmung und Digitalisierung.

Eng verbunden mit dem *intrinsischen Wert* ist der *Evidenzwert.*[3] Unter *Evidenzwert* ist hier die Bedeutung von Farblichkeit für die Interpretation von Archivgut zu verstehen, wie sie z.B. durch die Verwendung verschiedener Farben im Geschäftsgang bei Aktenschriftstücken oder durch Farben bei Wappendarstellung vorkommt.

Dem *Informationswert*[4] von Archivgut wird die Farbdigitalisierung insofern gerecht, als die vollständige Wiedergabe der Inhalte des Objektes wie z.B. Bildinhalte oft nur durch eine genaue Reproduktion der vorhandenen Farben erreicht wird. Dabei ist aber wie bei der bitonalen Digitalisierung und der Digitalisierung in Graustufen wegen der Masse der Archivalien – von Objekten mit besonderem intrinsischem Wert einmal abgesehen – nicht die ästhetische Qualität der Reproduktion ausschlaggebend.

Eine möglichst exakte Farbwiedergabe ist daher neben anderen Qualitätskriterien wie Schärfe oder die Vermeidung von Bildstörungen ein wichtiger Faktor bei der Digitalisierung von Archivgut. Um sie zu erreichen, ist vor allem die Anwendung erprobter, normgerechter Bearbeitungsverfahren – beispielsweise in einem Workflow von der Mikroverfilmung über die Filmdigitalisierung bis hin zur Bildbearbeitung – notwendig. Eine wichtige Bedeutung innerhalb eines solchen Workflows kommt dabei dem Einsatz eines sogenannten Colormanagement-Systems zu.

Das Grundproblem bei der Farbdigitalisierung und der anschließenden Nutzung und Weiterverarbeitung von Farbdigitalisaten besteht darin, daß dabei zum einen mit unterschiedlichen Geräten gearbeitet wird und diese zum anderen auch noch unterschiedliche Farbräume, d.h. Modelle, in denen Farben mathematisch beschrieben werden, unterstützen.[5] Dabei ist die Gefahr von Farbverlusten oder Farbverfälschung groß.

Um die Unterschiede bei der Farbausgabe zu berücksichtigen ist eine Art Dolmetscher erforderlich, der die Eigenheiten der einzelnen Geräte kennt. Ein solcher Dolmetscher kann als *Colormanagement* bezeichnet werden. Ein Colormanagement-System (CMS) sorgt für einheitliche Farben über Ein- und Ausgabegeräte hinweg, indem die Farbräume der einzelnen Geräte aufeinander abgeglichen werden, um jederzeit verläßliche, wiederholbare und vor al-

---

[3] Die Aussagekraft von Unterlagen über Abläufe und Verfahren in der Ursprungsstelle anhand von nicht geplanten formalen Merkmalen sowie von Verfügungen und Vermerken (Angaben), die im Entscheidungsprozeß für seine Steuerung entstanden sind. Siehe dazu Angelika *Menne-Haritz:* Schlüsselbegriffe der Archivterminologie. Lehrmaterialien für das Fach Archivwissenschaft (Veröffentlichungen der Archivschule Marburg 20). Marburg 1992. S. 47 und *http://staff-www.uni-marburg.de/~mennehar/germanterms.htm,* Art. Evidenzwert.

[4] Inhaltlicher Aussagewert von Archivgut über Fakten zu Personen, Orten und Ereignissen, die Gegenstand von Verwaltungshandeln geworden waren. Siehe dazu *Menne-Haritz*, Schlüsselbegriffe, S. 49 und *http://staff-www.uni–marburg.de/~mennehar/germanterms.htm*, Art. Informationswert.

[5] Zum Thema *Farbräume* siehe unten S. 184.

lem vorhersehbare Farbreproduktionen zu erzielen. Dabei wird für jedes beteiligte Gerät ein sogenanntes Farbprofil erstellt, das die Umrechnung zwischen den Farben dieses Gerätes und einem geräteunabhängigen, physikalisch definierten Farbraum erlaubt.[6] Ein weiteres Ziel von Colormanagement ist es, einen Ausgabeprozeß auf einem anderen Ausgabegerät zu simulieren wie z.B. eine Druckausgabe auf einem Monitor (Softproof).

Um all dies zu gewährleisten müssen Standards definiert werden, die den gesamten Produktionsprozeß berücksichtigen. Neben den technischen Unterschieden müssen die verschiedenen Methoden der Farberzeugung und die unterschiedlichen Farbräume der beteiligten Geräte berücksichtigt werden.

Bevor das Thema *Farbmanagement* und seine Komponenten näher beschrieben wird, soll zunächst allgemein auf *Farbe* und verschiedene Farbmodelle eingegangen werden.

Wahrnehmung und Beschreibung von Farben

Der Mensch nimmt *Farbe* in Form bestimmter Wellenlängen des Lichts wahr.[7] Zerlegt man Licht in seine verschiedenen Wellenlängen entsteht ein Spektrum mit allen möglichen Farben. Dabei ist Farbe eine Sache der Empfindung, der subjektiven Wahrnehmung. Beim Betrachten eines Gegenstandes nehmen verschiedene Personen die genau gleiche Farbe unterschiedlich wahr.

Der Farbeindruck wird von vielen Faktoren beeinflußt.[8]

1. *von der Art der Lichtquelle*: Bei unterschiedlicher Beleuchtung, z.B. durch Sonnenlicht, Leuchtstoffröhren oder Glühlampen, sieht ein Objekt jedesmal anders aus.
2. *vom individuellen Farbempfinden*: Bei jedem Menschen ist die spektrale Empfindlichkeit des Auges etwas anders.
3. *von der Objektgröße*: Farben, die eine große Fläche bedecken, wirken in der Regel leuchtender und intensiver als auf kleiner Fläche.

---

6 Zur Definition von *Colormanagement* siehe Jan-Peter *Homann:* Digitales Colormanagement. Farbe in der Publishing-Praxis (Edition Page). Berlin u.a. 1998. S. 65. – Rudolph E. *Burger:* Colormanagement. Konzepte, Systeme, Begriffe. Übersetzung von Thomas *Merz* (Edition Page). Berlin u.a. 1997. S. 25. – Die Geheimnisse des Farbmanagements. Die Digitale Farbe, Teil 5. Hg. von *Agfa-Gevaert* N.V., Mortsel (Belgien) 1997. S. 16. – Rüdiger *Maaß:* Farbe bekennen – Colormanagement in Theorie und Praxis, Teil 1. In: Print and Produktion 11/98 S. 38f. und Color Management. Die kreative Welt der digitalen Daten. Hg. von Linotype CPS, Bad Homburg o.J (PDF-Datei enthalten auf CD *Heidelberg. Color Publishing Solutions Version 4*), Art. Das Ziel.

7 Zur Farbwahrnehmung vgl. Color Management, wie Anm. 6 Kap. Farbempfinden. – *Homann,* Digitales Colormanagement, wie Anm. 6, S. 8–11.

8 Vgl. Exakte Farbkommunikation. Vom Farbgefühl zur objektiven Messung. Broschüre der Minolta GmbH. Ahrensburg 1996. S. 7.

4. *vom Objekthintergrund*: Befindet sich ein Objekt vor einem hellen Hintergrund, wirkt seine Farbe dumpfer als vor einem dunklen Hintergrund.
5. *von der Betrachtungsposition*: Die Farbe eines Objektes wirkt verschieden hell oder intensiv, je nachdem aus welchem Winkel es betrachtet wird. Der Farbeindruck ist also richtungsabhängig.

Das Problem der unterschiedlichen Farbwahrnehmung kann nur über standardisierte Methoden zur Bezeichnung von Farben in Form einer mathematischen Beschreibung gelöst werden. Dabei müssen standardisierte Bedingungen für eine objektive Farbbeschreibung geschaffen werden.[9]

Zur besseren Beschreibung von Farbunterschieden klassifiziert man jede Farbe nach folgenden Merkmalen:[10]

1. nach dem Farbton:

Die uns bekannten Namen für Farben wie rot, gelb, blau bezeichnen die Farbtöne. Der Farbton ist die Eigenschaft einer Farbe, die durch die Wellenlängen des von einem Gegenstand reflektierten Lichts bestimmt wird. Neue Farbtöne ergeben sich, wenn man unterschiedliche Farben mischt. Aus roter und gelber Farbe entsteht z.B. orange. Reiht man alle Farbtöne aneinander, vereinen sie sich zu einem sogenannten Farbkreis. Im Hinblick auf Farbraummodelle ist der Farbton ein Punkt im Farbraum, der eine Farbe repräsentiert.

Abb. 1: Farbkreis.

[9] Vgl. Color Management, wie Anm. 6, Kap. Farbmessung.
[10] Vgl. Die Geheimnisse des Farbmanagements, wie Anm. 6, S. 8f. – Minolta, Exakte Farbkommunikation, wie Anm. 8, S. 10f.

2. nach der Helligkeit:

Die Helligkeit, auch als Leuchtkraft bezeichnet, gibt an wie hell oder dunkel eine Farbe ist. Die Helligkeit einer Farbe kann unabhängig von ihrem Farbton bestimmt und gemessen werden.

3. nach der Sättigung:

Unter Sättigung versteht man die Farbkraft, Farbintensität, Farbreinheit oder Stärke einer Farbe und ihre Abweichung von Grau. Eine Farbe ist gesättigt, wenn sie nicht stumpf oder grau wirkt. Die Sättigung ist unabhängig vom Farbton und der Helligkeit.

Mit den drei Basiseigenschaften Farbton, Helligkeit und Sättigung läßt sich *Farbe* dreidimensional in Form eines sogenannten Farbkörpers darstellen.

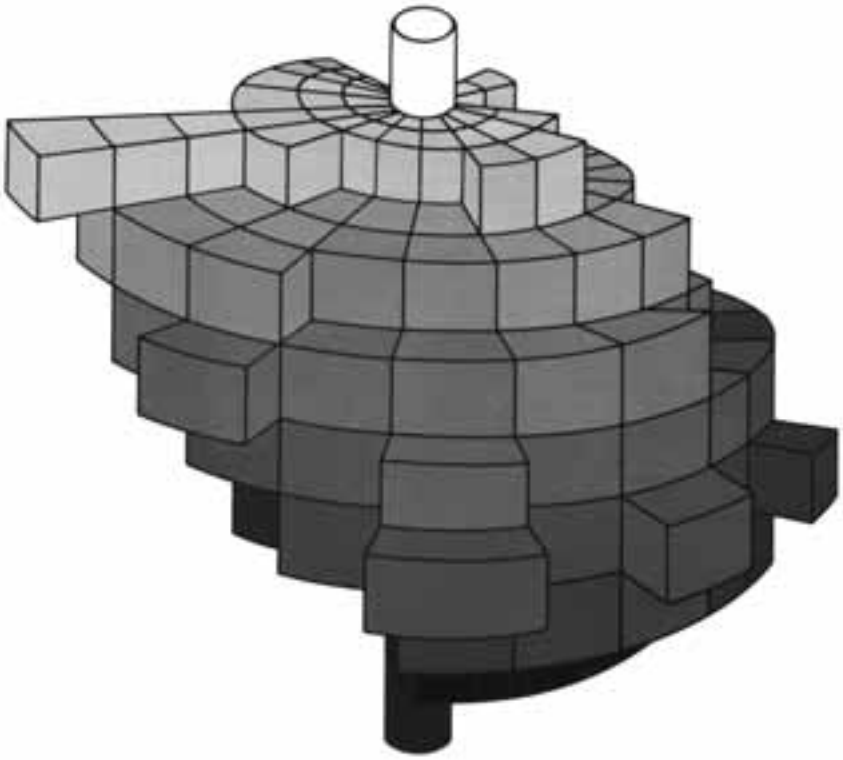


Abb. 2: Dreidimensionaler Farbkörper.

Die Farbtöne liegen beim dreidimensionalen Farbkörper auf dem Außenmantel, die Helligkeit verändert sich auf der senkrechten Achse und die Sättigung ist außen am höchsten und verringert sich zur Mittelachse hin. Da der horizontal verlaufende Sättigungsgrad mit den jeweiligen Farbtönen und Helligkeiten variiert, ergibt sich ein recht unregelmäßiges Gebilde.

Es gibt zwei Methoden zur Erzielung einer bestimmten Farbe.[11] Die eine Methode basiert auf Licht, die andere auf Druckfarben. Die erste wird als additive Farbmischung bezeichnet, die zweite als subtraktive Farbmischung. Die additive Farbmischung wird u.a. bei Farbmonitoren angewandt. Durch Aussendung von rotem, grünem und blauem Licht in unterschiedlicher Intensität können Monitore alle anderen Farben darstellen. Durch die Mischung der drei Primärfarben entsteht weißes Licht. Im Bereich der Drucktechnik kommt dagegen die subtraktive Farbmischung zum Einsatz. Die drei Primärfarben beim Druck sind Cyan, Magenta und Gelb. Auf einer weißen Oberfläche z.B. Papier werden Druckfarben verwendet, die die unerwünschten Farben ausfiltern oder subtrahieren. Eine Druckfarbe subtrahiert das Licht aller Farben, mit Ausnahme der eigenen Farbe. Werden alle drei Primärfarben übereinander auf einer Seite ausgebracht, wird das gesamte Licht absorbiert und die Seite erscheint schwarz.

Farbraumtheorien und -modelle

Zur exakten mathematischen Definition von Farben dienen Farbmaßsysteme oder Farbraummodelle.[12] Sie erlauben die Kennzeichnung der Farbe eines Objektes mit Hilfe von Zahlenwerten und dienen der modernen Farbkommunikation. So muß z.B. jedes Colormanagement auf einer Referenz aufsetzen, d.h. auf colormetrisch genau definierten Daten eines möglichst umfassenden und für die digitale Bildverarbeitung geeigneten geräteunabhängigen Farbraums.[13]

Ein Farbraum ist ein Modell um Farbe, wie sie vom menschlichen Auge wahrgenommen wird, quantisierbar zu machen, d.h. mathematisch in Zahlenform darzustellen.[14] Bei Farbräumen muß zwischen geräteabhängigen und geräteunabhängigen Farbräumen unterschieden werden.

---

[11] Vgl. Die Geheimnisse des Farbmanagements, wie Anm. 6, S. 7. – *Homann*, Digitales Colormanagementwie Anm. 6, S. 12–15.

[12] Einen Überblick über die verschiedenen Farbraummodelle siehe in Minolta, Exakte Farbkommunikation, wie Anm. 8, S. 14–20 und *Burger*, Colormanagement, wie Anm. 6, S. 6–10 und 17–22.

[13] Vgl. Helmut *Hofmann*: Haltbarkeit digitaler Bilder: Die Fakten. In: Publishing Praxis 11 (1997) S. 38–40, hier S. 38f.

[14] Siehe dazu LaserSoft SilverFast Ai Handbuch. Hg. von der LaserSoft Imaging Entwicklungs- und Vertriebs GmbH. Kiel 1999. S. 167.

Geräteunabhängige Farbräume

In geräteunabhängigen Farbräumen werden alle vom menschlichen Auge unterscheidbaren Farben dargestellt. Es erfolgt also kein Informationsverlust.

Die bekanntesten geräteunabhängigen Farbräume wurden von der Internationalen Beleuchtungskommission (CIE= Commission Internationale de l'Eclairage) entwickelt: das YxY-Farbsystem (1931), das CIE-xyz-System (1976) und das CIE L*a*b*-System (1976).[15]

Der L*a*b*-Farbraum, auch CIELAB-Farbraum genannt, bietet als sogenanntes gleichabständiges Farbmodell im Gegensatz zum YxY-Farbraum eine bessere Übereinstimmung zwischen geometrischem und empfundenem Farbabstand.[16] Daher ist er heute der gebräuchliche Farbraum für die Farbmessung und hat in vielen Anwendungsbereichen eine große Verbreitung gefunden. Der Farbraum des L*a*b*-Systems ist durch die Helligkeit L* und die Farbkoordinaten a* und b* gekennzeichnet.

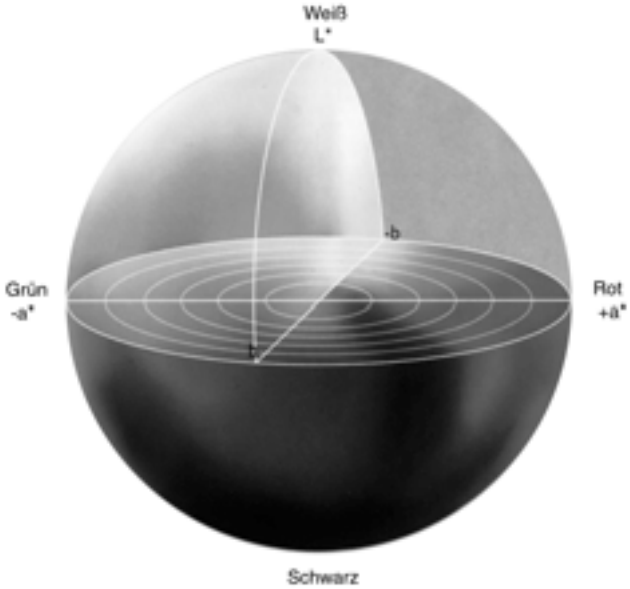


Abb. 3: Dreidimensionale Darstellung des L*a*b*-Farbraums: Die a-Achse erstreckt sich von Grün (–a) nach Rot (+a), die b-Achse von Blau (–b) nach Gelb (+b). Die Helligkeit (L) nimmt in diesem dreidimensionalen Körper von unten nach oben zu.

---

15 Vgl. Minolta, Exakte Farbkommunikation, wie Anm. 8, S. 14.

16 Zum L*a*b*-Farbraum siehe Color Management, wie Anm. 6 Art. Farbmessung. – *Homann*, wie Anm. 6, S. 34–41. – Die Geheimnisse des Farbmanagements, wie Anm. 6, S. 9.

Ein weiterer, relativ neuer geräteunabhängiger Farbraum ist der 1990 von Kodak und Philips geschaffene YCC-Farbraum.[17] Er ist wie die CIE-Farbräume colormetrisch definiert. Auf YCC können ebenfalls wie bei den CIE-Farbräumen sogenannte ICC-Profile[18] aufsetzen, wie sie von Farbmanagementsystemen verwendet werden. Der YCC-Farbraum ist momentan mit dem ebenfalls von Kodak entwickelten Imagepac-Dateiformat und dem für Online-Anwendungen entwickelten Flashpix-Format verbunden.[19] Da dieser Farbraum momentan noch ein proprietäres Farbmodell darstellt, sollten im Rahmen eines Colormanagement-Systems die standardisierten CIE-Farbräume, insbesondere der CIELAB-Farbraum, trotz mancher Nachteile gegenüber dem YCC-Farbraum bevorzugt Verwendung finden.

Geräteabhängige Farbräume

Scanner und Monitore bedienen sich eines geräteabhängigen RGB-Farbraums, dem als Primärfarben die Farben rot (**R**ed), grün (**G**reen) und blau (**B**lue) zugrunde liegen. *Den* genormten RGB-Farbraum gibt es nicht.[20] Zwar arbeitet jedes Gerät numerisch mit den gleichen RGB-Werten, *versteht* darunter aber eine andere Farbwahrnehmung. Erste Ansätze zu einer Normierung von RGB-Farbräumen bietet der u.a. von Hewlett-Packard und Microsoft propagierte sRGB-Farbraum, der rot, grün und blau als die Farben eines genau definierten theoretischen Monitors festlegt.[21] Der sRGB-Farbraum könnte zukünftig für eine gewisse Standardisierung im Bereich der Farbdarstellung in Online-Medien wie dem Internet sorgen und damit ein wichtiger Schritt bei Bemühungen um sog. Web-sichere Farben sein.

Der professionelle Vierfarbdruck bedient sich eines gerätespezifischen CMYK-Farbraums. Cyan, Magenta, Gelb (Yellow) und Schwarz (Black) sind die vier im normalen Farbdruck verwendeten Farben.[22]

Geräteabhängige Farbräume wie RGB oder CMYK können nur ca. 50% der im geräteunabhängigen CIELAB-Farbraum unterscheidbaren Farben abbilden. Um trotzdem eine realistische Wiedergabe zu erreichen, werden Farbtöne in ihren Abstufungen auf den Farbumfang des verwendeten Gerätes oder Prozesses reduziert. So ist eine absolut originalgetreue Reproduktion in den

---

17 Siehe dazu *Hofmann*, Haltbarkeit digitaler Bilder, wie Anm. 13, S. 39.

18 Sie dazu unten S. 189.

19 Zum Imagepac- und Flashpix-Dateiformat siehe unten S. 196.

20 Vgl. dazu auch Jörn *Loviscach*, Volles Spektrum. Farben verläßlich scannen, anzeigen und drucken. In: c't Magazin für Computertechnik 7 (1999) S. 122–127, hier S. 122.

21 Zum sRGB-Farbraum siehe *http://www.srgb.com* und *http://www.w3.org/Graphics/Color/sRGB*. – Vgl. auch *Loviscach*, Volles Spektrum, wie Anm. 20, S. 123.

22 Zum Vierfarbdruck und den CYMK-Farbräumen siehe z.B. *Homann*, Digitales Colormanagement, wie Anm. 6, S. 78ff.

Fällen nicht möglich, in denen Farbumfang und/oder Kontrastumfang des Originals größer ist als der der verwendeten Eingabe- und Ausgabegeräte. Geräteabhängige Farbräume eigenen sich daher nicht für digitale Masterformen, sondern nur für Anwendungformen.[23] Sie sind insbesondere für die Bildschirmausgabe und den Farbdruck erste Wahl. Dabei muß allerdings im Rahmen der noch vorzustellenden Farbraumtransformation eine Umwandlung des digitalen Masters von einem geräteunabhängigen Farbraum in einen geräteabhängigen Farbraum wie RGB für die Bildschirmdarstellung oder CMYK für den Druck vorgenommen werden.

Farbvermessung mit Meßgeräten unter Normbedingungen

Während das menschliche Auge für das genaue Bemessen von Farben nicht geeignet, ist dies mit einem Farbmeßgerät möglich. Ein solches Farbmeßgerät liefert numerische Werte der gebräuchlichen Farbraumstandards.

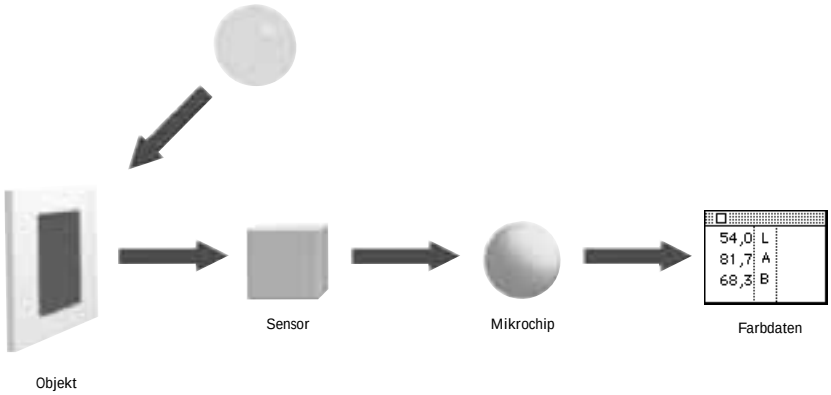


Abb. 4: Ablauf einer Farbmessung mit einem Spektralphotometer.

Farbmeßgeräte

Die Meßgeräte lassen sich dabei in zwei größere Gruppen unterteilen:[24] 1. Farbmeßgeräte nach dem Dreibereichsverfahren (Kolorimeter, Farbdensitometer) und 2. Farbmeßgeräte nach dem Spektralverfahren (Spektralphotometer, Photospektrometer).

---

23 Siehe dazu unten S. 196.

24 Vgl. Minolta, Exakte Farbkommunikation, wie Anm.8, S. 26–49. – *Burger*, Colormanagement, wie Anm. 6, S. 10–12.

Das Dreibereichsverfahren benutzt zur Messung des vom Objekt reflektierten Lichts drei Sensoren, die so gefiltert sind, daß sie den Empfindlichkeitskurven des menschlichen Auges entsprechen. Dabei reagieren sowohl das Auge als auch die Meßgerätsensoren auf die drei sogenannten Primärfarben rot, grün, blau.

Bei der spektralphotometrischen Methode wird das Licht in seinen Längenwellenanteile zerlegt, die anschließend mit vielen Lichtsensoren gemessen werden. Da Spektralphotometer absolute Farbdaten liefern, bieten sie eine höhere Genauigkeit als Dreibereichsmeßgeräte.

Nach der Abtastung erfolgt die Umrechnung der Meßwerte in die wirklichen Farbwerte. Oft bieten die Geräte dabei die Möglichkeit, Farbwerte für verschiedene geräteunabhängige Farbräume auszugeben.

Normlichtarten

Da, wie schon erwähnt, die Art der Beleuchtung den Farbeindruck beeinflußt, hat die CIE für die Farbmessung verschiedene Normlichtarten festgelegt.[25]

1. *Normlichtart D65*: entspricht mittlerem Tageslicht (einschließlich UV-Bereich) mit einer Farbtemperatur von ca. 6 500 Kelvin und empfiehlt sich daher für Messungen von Objekten unter Tageslichtbedingungen einschließlich UV-Strahlung.
2. *Normlichtart C*: entspricht mittlerem Tageslicht (ohne UV-Bereich) mit einer Farbtemperatur von ca. 6 800 Kelvin und eignet sich für Messungen von Objekten unter Tageslicht des sichtbaren Spektrums ohne UV-Strahlung.
3. *Normlichtart A*: entspricht Glühlampenlicht mit einer Farbtemperatur von ca. 2 850 Kelvin und ist für Messungen von Objekten unter künstlicher Glühlampenbeleuchtung geeignet.

Entscheidend für eine korrekte Farbmessung sind gleichbleibende Beleuchtungsbedingungen, d.h. eine gleichbleibende Lichtquelle des Meßgeräts und der Umgebungsbeleuchtung. Für Transmissionsmessungen von Filmmaterial sind oft zusätzliche Leuchttische mit Normlicht notwendig.

Es ist festzuhalten, daß für die exakte Reproduzierbarkeit von Farben der Einsatz eines Farbmeßgerätes in Kombination mit anderen Colormanagement-Systemkomponenten eine unabdingbare Voraussetzung darstellt. Dabei spielen Farbmeßgeräte besonders für die Erstellung von Geräteprofilen eine wichtige Rolle.

---

[25] Siehe dazu Minolta, Exakte Farbkommunikation, wie Anm. 8, S. 40f.

Die Erstellung von Geräteprofilen bei der Charakterisierung und Kalibrierung von Ein- und Ausgabegeräten[26]

Ein CMS benötigt neben der Qualifizierung der Anwender mindestens zwei technische Voraussetzungen: 1. Profile standardisierter Prozesse und Geräte einschließlich der dafür notwendigen Meßgeräte, Testvorlagen und Profilgenerierungs-Software und 2. Software für Farbtransformationen, die in der Lage ist, einzelne Profile miteinander zu verknüpfen, um einen konstanten Farbabstand über den gesamten Prozeßablauf zu gewährleisten. Angestrebt werden muß dabei, daß die Farbraumtransformation nicht selbständig von den jeweiligen Anwendungsprogrammen durchgeführt wird, sondern alle Programme das gleiche Umrechnungsmodul auf der Ebene des Betriebssystems nutzen.

Farbprofile oder Geräteprofile, die die Basis für farbverbindliches Arbeiten in den Anwendungsprogrammen und auf Systemebene bilden, ergeben sich durch eine Beschreibung aller in die Prozeßschritte der Digitalisierung und Bildbearbeitung eingebundenen Ein- und Ausgabegeräte. Erstellt werden sie im Zuge der sogenannten Gerätecharakterisierung und -kalibrierung. Hierbei handelt es sich um einen Vorgang, bei dem der Zusammenhang zwischen dem eigenen Farbraum eines Geräts und einem geräteunabhängigen Referenzfarbraum definiert wird. Unter Gerätecharakterisierung versteht man die Grundjustierung der Farbbalance eines Gerätes beim Hersteller. Diese kann sich im Laufe der Zeit verändern (z.B. durch Veränderung der Farbtemperatur einer Scannerlampe), so daß man durch die sogenannte Gerätekalibrierung (auch Linearisierung genannt) für ein Ausgleichen der Farbbalance eines Geräts sorgen muß.[27] Durch die Kalibrierung erreicht man, das sich ein Gerät gemäß der Parameter verhält, die bei seiner Charakterisierung festgelegt wurden, es also ein gleichbleibendes Farbverhalten aufweist. Sie erfolgt durch den Benutzter mittels Testvorlagen. Alle Ein- und Ausgabegeräte innerhalb eines CMS müssen vor ihrer Verwendung charakterisiert und kalibriert werden.

Als Farbprofile von Geräten sollten sogenannte ICC-Profile erzeugt werden, da diese sich als Standard etabliert haben.[28] Dabei beziehen sich die Gerä-

---

26 Vgl. dazu Color Management, wie Anm. 6, Art. Die Bedingungen.

27 Zur Gerätekalibrierung vgl. Jan-Peter *Homann*, Color Management. Richtig digital proofen, in: Page Prepress 06/97, S. 18–21, hier S. 18.

28 Zur Schaffung einheitlicher Standards auf dem Gebiet der Farbdatenübertragung haben sich 1992/93 verschiedene Hersteller (u.a. Adobe, Agfa, Apple, Kodak, Linotype-Hell, Microsoft, Sun) zum sog. Internationalen Color Consortium (ICC) zusammengeschlossen. Siehe dazu *http://www.color.org* und *Homann*, Digitales Colormanagement, wie Anm. 6, S. 98–105. ICC-Profile haben einen einheitlichen Aufbau. Zur Entschlüsselung von Profildateien gibt es spezielle Software wie das Programm ICCread (erhältlich unter *ftp://www.heise.de/ct/ftp/*).

teprofile auf die geräteunabhängigen CIE-Farbmodelle. Die Erstellung von ICC-Profilen soll für Eingabegeräte am Beispiel eines Flachbettscanners und für Ausgabegeräte am Beispiel eines Farbmonitors beschrieben werden.

Profilgenerierung bei Scannern[29]

Das Charakterisieren und Kalibrieren des Scanners ist der erste wichtige Schritt bei der Einrichtung eines Farbmanagementsystems, um die kleinen Farbveränderungen zu ermitteln, die das betreffende Gerät jedesmal einbringt, wenn ein Bild gescannt wird. Dabei wird der geräteeigene Farbraum eines Scanners, der sogenannte Gamut, ermittelt. Im Normalfall handelt es sich hier um einen geräteabhängigen RGB-Farbraum, so daß verschiedene Scanner die gleiche Farbe unterschiedlich *sehen* aufgrund unterschiedlicher Eigenschaften ihrer Abtastsysteme.

Die Kalibrierung muß möglichst unter optimalen Bedingungen erfolgen. Eine Referenzvorlage mit exakt definierten Farbfeldern wird mit dem Scanner eingescannt.[30] Dabei wird eine spezielle Kalibrierungssoftware mit der Möglichkeit der Profilgenerierung verwendet. Die *tatsächlich* gemessenen Farbwerte werden mit den *idealen* Werten in Beziehung gesetzt. Die Idealwerte werden vom Hersteller der Referenzkarte auf Diskette geliefert oder mittels eines Spektralphotometers durch Vermessung der Testvorlagen ermittelt. Anhand dieser beiden Datensätze wird mittels der Profilgenerierungssoftware ein sogenanntes Geräteprofil erstellt, das angibt wo der Scanner von den Idealwerten abweicht.

Sobald das Farbmanagementsystem die individuellen Parameter des kalibrierten Scanners kennt, können diese bei dem Scanvorgang berücksichtigt werden. Liefert ein Scanner z.B. Bilder mit etwas zu starken Blautönen und zu schwachen Rottönen, kann das Farbmanagementsystem anhand des Geräteprofils die Blau- und Rottöne entsprechend korrigieren.

---

29 Vgl. *Homann*, Digitales Colormanagement, wie Anm. 6, S. 68 und S. 127–135.

30 Für Flachbettscanner wird als Referenzvorlage meist eine sog. ANSI IT8.7-Farbkarte verwendet. Für ihre Herstellung werden Farbmuster aus dem CIELAB-Farbraum verwendet. Mit der ANSI IT8.7/1–1993-Farbkarte können Farbwerte für Durchsichtvorlagen, mit der ANSI IT8.7/2–1993-Farbkarte solche für Aufsichtsvorlagen gemessen werden. Produziert werden solche Farbkarten von den Firmen Agfa, Fuji und Kodak.

Profilgenerierung bei Farbmonitoren[31]

Eine exakte Farbreproduktion auf einem Monitor ist nicht leicht zu erreichen. Jeder Monitor zeigt Farben anders an. Dabei kann ein Monitor nur Farben innerhalb seines RGB-Gamuts anzeigen. Neben unterschiedlichen Fertigungsspezifikationen spielt auch die Temperatur und Alterung des jeweiligen Geräts eine Rolle. Nach mehreren Betriebsstunden kann ein und derselbe Monitor Farben eventuell völlig anders darstellen. Auch die Umgebungsbeleuchtung am Arbeitsplatz beeinflußt die Farbdarstellung. Externe Lichtquellen vermischen sich mit den Farben auf dem Monitor, wenn ihr Licht auf die Außenfläche des Bildschirms trifft. Außerdem verändert sich die Farbwiedergabe bezüglich Helligkeit und Sättigung durch manuelles Einstellen von Helligkeit und Kontrast. Ein regelmäßiges Kalibrieren des Monitors – am besten bei jeder Veränderung der Umgebungsbedingungen ist daher unerläßlich für eine optimale Farbtreue.

Für eine komplette Kalibrierung müssen vier Parameter eingestellt werden: Helligkeit, Kontrast[32], Gamma[33] und Weißpunkttemperatur[34]. Helligkeit und Kontrast werden mit Hilfe von Reglern am Monitor manuell eingestellt. Gamma und Weißpunkt werden in der Systemsoftware bzw. in der Anwendungssoftware eingestellt. Im nächsten Schritt wird mit einem Farbmeßgerät, das mit einem Saugnapf am Monitor befestigt wird, und entsprechender Software, ein Geräteprofil des jeweiligen Monitors erzeugt.[35] Die Profilerzeugungssoftware übernimmt dabei die Steuerung des Monitors und weist diesen an, verschiedene Farben darzustellen, die dann gemessen werden. Die Software erstellt ein Geräteprofil, indem die tatsächlichen, gemessenen Farbwerte mit den idealen Farbwerten, die erwartet werden, in Beziehung gesetzt werden.

---

31 Vgl. Die Geheimnisse des Farbmanagements, wie Anm. 6, S. 20f. – Ulrich *Schurr:* Bildschirm-Abgleich. In: Profi Publisher 3/97 S. 32f. – Color Management, wie Anm. 6, Art. Von der Vorlage zum Monitorbild. – *Homann*, Digitales Colormanagement, wie Anm. 6, S. 69.

32 Das Verhältnis zwischen den hellsten und dunkelsten Flächen eines Bildes bezeichnet man als *Kontrast*.

33 Unter *Gamma* versteht man das Verhältnis zwischen den Tonwerten in einer Bilddatei zu den Tonwerten, die von einem Monitor oder Ausgabegerät erzeugt werden.

34 Der *Weißpunkt* ist ein veränderlicher Referenzpunkt, der den hellsten Bereich in einem Bild definiert, und nach dem alle anderen Bereiche eines Bildes eingestellt werden können.

35 Vgl. *Homann*, Digitales Colormanagement, wie Anm. 6, S. 148.

Profilgenerierung bei Druckausgabegeräten[36]

Auf die Charakterisierung und Kalibrierung der verschiedenen Druckausgabegeräte kann hier nur kurz eingegangen werden. Sie ist vergleichbar mit dem entsprechenden Verfahren für einen Scanner oder Monitor. Für den Druck findet normalerweise eine Farbraumtransformation von einem geräteunabhängigen CIE-Farbraum in einen geräteabhängigen CMYK-Farbraum statt. Der Vorgang der Profilgenerierung sieht vereinfacht folgendermaßen aus: Zunächst wird eine Referenz-Farbkarte,[37] die als Druckdatei vorliegt auf dem entsprechenden Gerät ausgegeben. Anschließend werden die Farbfelder der ausgedruckten Farbkarte mit einem Farbmeßgerät (Kolorimeter oder Spektralphotometer) vermessen. Nach dem Messen werden die tatsächlichen Meßwerte mit den idealen Farbwerten der Referenzkarte durch entsprechende Profilgenerierungssoftware verglichen und ein Geräteprofil erstellt.

Bei Druckausgabegeräten ist zu beachten, daß für die Profilgenerierung die verwendeten Druckmedien wie Papier oder Druckmaterialien wie Farbe einen Einfluß auf die Profilgenerierung haben. Für die Kalibrierung bedeutet die Abhängigkeit vom Trägermaterial, daß für jede Papiersorte ein eigenes Geräteprofil erstellt werden sollte.

Die Arbeit mit den Profilen bei der Farbraumtransformation

Nachdem für alle in einem CMS-Workflow beteiligten Ein- und Ausgabegeräte Geräteprofile erstellt worden sind, können diese für Farbraumtransformationen verwendet werden.

Eine besondere Herausforderung für ein CMS ist die Umrechnung des Kommunikationsfarbraum in den Farbraum des Ausgabegeräts. Dabei muß berücksichtigt werden, daß beispielsweise Monitore nicht alle Farben darstellen können, die das menschliche Auge wahrnehmen kann, ein Drucker wiederum nicht alle Farben drucken kann, die ein Monitor darstellen kann. Dieses Problem kann nur über eine möglichst geschickte Farbraumanpassung, das sogenannte Gamut Mapping,[38] vorgenommen werden. Hierbei werden die Farbräume von der Eingabe über alle Zwischenschritte bis zur Ausgabe so aufeinander abgestimmt, daß die nicht darstellbaren Farben sinnvoll durch an-

---

36 Vgl. dazu Die Geheimnisse des Farbmanagements, wie Anm. 6, S. 24f. – Color Management, wie Anm. 6, Art. Von der Vorlage zum Farbausdruck. – *Homann*, Digitales Colormanagement, wie Anm. 6, S. 70f.

37 Hier wird meist die sog. ANSI IT8.7/3-Farbkarte verwendet.

38 Siehe dazu Color Management, wie Anm. 6, Art. Der Weg. – *Burger*, Colormanagement, wie Anm. 6, S. 26.

dere Farben ersetzt werden. Ziel ist hierbei, daß der Gesamteindruck der Vorlage, also das Verhältnis der Farben zueinander, erhalten bleibt.

Für Farbraumtransformationen haben CMS drei Möglichkeiten:[39]

1. *Wahrnehmungsbasierte Anpassung*: Alle Farben, auch die von dem jeweiligen Ausgabegerät darstellbaren, werden so weit nach Grau verschoben, bis auch die Extreme darstellbar oder druckbar sind. Der Gesamteindruck von Bildern bleibt dabei erhalten, Grafiken werden aber flau.
2. *Sättigungsbasierte Anpassung*: Jede nicht darstellbare Farbe wird durch eine ähnliche, darstellbare ersetzt. Auf diese Weise bleibt die Darstellung kräftig, allerdings wirken Bilder oft in zu grellen Tönen.
3. *Farbmetrische Anpassung*: Jede nicht darstellbare Farbe wird durch die der menschlichen Empfindung nächstliegende darstellbare Farbe ersetzt.

Damit auf einem anderen System beispielsweise ein Bild so dargestellt wird wie auf dem lokalen, benötigt dieses nicht nur die Bilddaten sondern auch das Monitorprofil, um die Farbdarstellung anzugleichen.

Dabei kann das erstellte Geräteprofil direkt auf dem Rechner verankert werden, so daß man über das CMS-Modul darauf zugreifen kann. Die beste systemweite CMS Unterstützung gibt es zur Zeit auf der MacOS-Plattform. Hier ist mit ColorSync ein Farbmanagementsystem im Betriebssystem integriert, auf das dann die einzelnen Anwendungsprogramme und ICC-Profile aufsetzen können.[40] Für die Microsoft Windows-Umgebung wurde von Microsoft der ICM 2.0-Standard geschaffen,[41] der aber momentan von Anwendungssoftware kaum unterstützt wird, so daß man unter dieser Plattform bisher auf bestimmte CMS-Funktionen innerhalb der Anwendungssoftware wie z.B. Adobe Photoshop 5.0 und zusätzlich spezieller Profilgenerierungssoftware angewiesen ist.

Neben einer Verankerung der Geräteprofile in einem CMS-Modul ist es auch möglich, ein Profil direkt bei den Bilddaten zu hinterlegen. Dies bietet sich an, wenn die Daten weitergegeben werden und das Ausgabemedium noch nicht bekannt ist. Sinnvoll funktioniert dies momentan nur beim TIF-Format.[42] Allerdings gibt es bislang noch kaum Software, die solche eingebetteten Profile auch auswerten kann.

Die Arbeit mit vielen einzelnen Profilen hat oftmals eine große Anzahl von Farbraumtransformationen zur Folge, was zeitaufwendig ist. Außerdem wür-

---

39 Siehe dazu *Loviscach*, Volles Spektrum, wie Anm. 20, S. 122f.

40 Zu ColorSync für MacOS siehe *Homann*, Digitales Colormanagement, wie Anm. 6, S. 136.

41 Zu ICM (Image Color Matching) siehe *http://www.microsoft.com*.

42 Siehe dazu unten S. 196.

den bei jeder zusätzlichen Transformation die Daten etwas stärker abweichen. Das Problem kann gelöst werden, wenn man mehrere Profile zu einer sogenannten Transformationstabelle oder einem Gesamtprofil verknüpft. Dadurch wird die Farbraumanpassung genauer und schneller als bei mehreren Einzeltransformationen hintereinander.[43]

Eine typische Farbraumtransformation verläuft folgendermaßen: Die geräteabhängigen Farbräume der Eingabegeräte (meist RGB) werden mittels der erstellten Profile in einen geräteunabhängigen Farbraum (meist CIE LAB) transformiert. Anschließend wird dieser dann wieder in einen geräteabhängigen Farbraum eines Ausgabegeräts transformiert. Für jede Farbumrechnung werden zwei ICC-Profile benötigt, zum einen das Quellprofil mit einer Beschreibung des Farbraums im Ursprungszustand, zum anderen das Zielprofil.

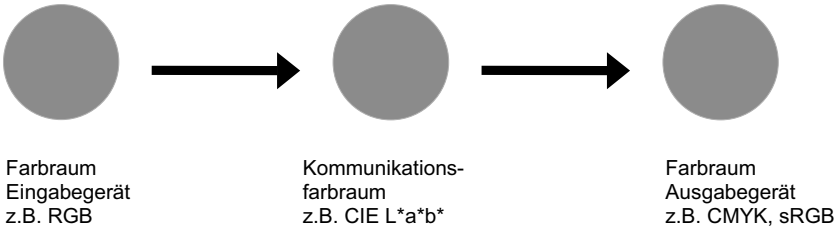


Abb. 5: Farbraumtransformation: Die spezifischen Daten des Eingabemediums werden in einen geräteunabhängigen Kommunikationsfarbraum transformiert und anschließend in den Farbraum des Ausgabegeräts umgerechnet.

Colormanagement bei der Farbdigitalisierung von Mikroformen

Eine besondere Herausforderung für ein Farbmanagement ist die Digitalisierung von Farbmikroformen. Hier gibt es im Vergleich zur Direktdigitalisierung durch den Film eine Zwischenstufe, die die Farbauthentizität der digitalen Master bezogen auf die Ausgangsobjekte schwieriger macht. Im Rahmen des Forschungsprojektes *Digitale Konversionsformen* der Landesarchivdirektion Baden-Württemberg werden erste Möglichkeiten eines Colormanagements bei der Digitalisierung von Farbmikroformen untersucht. Genauere Erkenntnisse können allerdings erst im Rahmen weiterer Untersuchungen gewonnen werden.

---

[43] Siehe dazu Color Management, wie Anm. 6, Art. Die Bedingungen.

Für die Digitalisierung vom Farbfilm bietet sich ein hochauflösender und farbbeständiger Color-Mikrofilm nach dem Farbbleichverfahren auf Polyester-Filmunterlage an. Hinsichtlich der Qualität sollte dabei immer vom Aufnahmefilm gescannt werden.[44] Die visuelle Qualitätskontrolle der von Dienstleistern erstellten filmdigitalisierten Formen hat gezeigt, daß für eine korrekte Ton- und Farbwiedergabe die Einführung eines Colormanagements in Verbindung mit einer Gerätecharakterisierung und -kalibrierung eine unabdingbare Voraussetzung ist. Dies gilt vor allem für die Digitalisierung von sequentiellen Dokumenten eines Objektes oder gleichartiger Objekte.

Colormanagement-Verfahren werden im Rahmen des Projekts bereits bei einem der Dienstleister für Farbfilmdigitalisierung in einer Apple-Macintosh-Umgebung erprobt. Dabei wird schon bei der Microverfilmung eine Testvorlage für eine Filmemulsion erstellt, die die Aufnahmebedingungen charakterisiert. Berücksichtigt werden dabei Vorlagenhalterung (mit/ohne Glas), die Beleuchtungssituation (Lampenart, Filterung, Ausleuchtung je nach Vorlagengröße) und das verwendete Objektiv (Vorlagengröße/Verkleinerungsfaktor).

Vor der Digitalisierung mit dem Kamera-Scanner wird für diesen ein ICC-Geräteprofil mit spezieller Profilgenerierungssoftware erstellt. Das Profil wird dann unabhängig vom Anwendungsprogramm im Betriebssystem (MacOS ColorSync) hinterlegt. Die Kalibrierung qualitativ hochwertiger, großer Monitore erfolgt mittels Meßgeräte. Auch hier werden anschließend ICC-Profile erstellt und im Betriebssystem hinterlegt. Dabei werden die Bildschirme blendfrei in Räumen mit farbneutralen Normlicht aufgestellt.

Die Bildbearbeitung erfolgt mit Adobe Photoshop 5.0, da diese Version die Einbindung und Auswertung von ICC-Profilen ermöglicht. Von seiten des Projektteams ist geplant, ein Farbmanagementsystem auch auf den Projekt-PCs zu installieren, so daß eine farbkonstante Auswertung und Weiterverarbeitung der von Dienstleistern gelieferten digitalen Formen möglich sein wird.

Momentan werden die Scandaten noch im RGB-Farbraum belassen (TIFF verlustfrei) und nicht in einen geräteunabhängigen Farbraum wie CIELAB konvertiert, da bisher diverse Anwendungsprogramme diese Farbräume noch nicht unterstützen und die Daten unter Zeitaufwand z.B. in einen RGB-Monitor-Farbraum konvertiert werden müssen. Langfristig wird aber die Verwendung geräteunabhängiger Farbräume für digitale Master angestrebt.

---

44 Siehe Hartmut *Weber* und Marianne *Dörr:* Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft. In: Zeitschrift für Bibliothekswesen und Bibliographie 44 (1997) Heft 1 S. 53–76, hier S. 60ff. (auch im Internet unter *http://www.lad-bw.de/lad/dfgdigh1.htm#Qualität* abrufbar).

Medienunabhängige Farbräume und Dateiformate für digitale Master

Für ein CMS-Workflow ist wie schon erwähnt die Wahl eines geräte- oder medienunabhängigen Farbraumes entscheidend.[45] Speicherung unter Verwendung von geräteabhängigen, nicht colormetrisch definierten Farbräumen wie RGB oder CMYK führen zum Verlust von Farbinformationen.

Um die digitalen Master auf Datenträger wie CD-R oder DVD zu speichern, muß zunächst eine Farbraumtransformation mit dem Profil des Eingabegeräts vorgenommen werden. Falls kein Geräteprofil vorliegt, sollte auf den Datenträger neben den gewünschten digitalen Bildern auch eine Testvorlage für die Profilgenerierung mit abgespeichert werden. Sie sollte mit dem gleichen Scanner wie die Ausgangsobjekte eingescannt werden. Mit der digitalisierten Testvorlage kann dann später durch eine Profilgenerierungssoftware ein Geräteprofil erstellt werden, das für die eingescannten Bilddaten verwendet werden kann.[46]

Aus Gründen der wirtschaftlichen Speicherung und praktikablen Verarbeitung der Bilddaten spielen bei der Digitalisierung in Farbe Kompressionsverfahren eine noch größere Rolle als bei bitonaler Digitalisierung oder Digitalisierung mit Graustufen. Dabei ist zu beachten, daß digitale Master nicht farbreduziert oder mit verlustbehafteten Kompressionstechnologien abgespeichert werden. Komprimierungsverfahren, wie sie beispielsweise beim JPEG-Dateiformat Verwendung finden, führen nicht nur zur Verschlechterung der Wiedergabequalität im Bereich der Bildschärfe sondern auch zu Farbverfälschungen.

Momentan ist das TIF-Format das einzige quasi plattformübergreifende und nicht proprietäre Standarddateiformat für *digitale Master*, das eine verlustfreie Komprimierung zuläßt und geräte- oder medienunabhängige Farbräume wie den CIELAB-Farbraum unterstützt.[47] Es könnte in naher Zukunft vom freien PNG-Format und vom proprietären Imagepac-Format[48]

---

45 Vgl. dazu auch *Hofmann*, Haltbarkeit digitaler Bilder, wie Anm. 13, S. 38f.

46 Vgl. dazu auch Color Management, wie Anm. 6, Art. Die Photo-CD.

47 Zu den Vor- und Nachteilen der verschiedenen Dateiformate und Komprimierungstechnologien siehe in diesem Band: Gerald *Maier:* Qualität, Bearbeitung und Präsentation digitaler Bilder. S. 129.

48 Das Imagpac-Dateiformat wurde von der Firma Kodak entwickelt und macht zusammen mit dem YCC-Farbraum den Kern des ebenfalls von Kodak entwickelten Photo-CD-Standards aus. Siehe dazu *Hofmann*, Haltbarkeit digitaler Bilder, wie Anm. 13, S. 39. – Vgl. auch *http://www.ImageManagement.de* und *http://www.photocd.de*. – Zum FlashPix-Dateiformat siehe auch Roland *Dreyer* und Helmut G. *Hofmann*: FlashPix – Postscript der PC-Fotographie. In: OP Druck Magazin 3 (1998) (auch als Sonderdruck erhältlich). – Philip *Greenspun:* Publishing FlashPix (*http://photo.net/photo/publishing-flashpix.html*) und *http://www.kodak.com/US/en/digital/flashPix/flashPixFAQ.shtml.*

Konkurrenz bekommen. Das PNG-Format bietet bisher noch keine Unterstützung des geräteunabhängigen CIELAB-Farbraums und läßt im Gegensatz zu TIF bisher auch nicht die Einbindung von ICC-Profilen zu. Das Imagepac-Format besitzt als Neuerung gegenüber dem TIF-Format eine sogenannte Multi-Resolution Fähigkeit, d.h. in einer Datei sind mehrere Auflösungen gespeichert, so daß ein Zoom-Effekt erreicht werden kann.

Die CMS-Prozeßkette für die Farbdigitalisierung von Archivgut

Der Workflow bzw. die Prozeßkette für die Farbdigitalisierung von Archivgut unter Einsatz von Colormanagement kann vereinfacht folgendermaßen aussehen:

1. Charakterisierung und Kalibrierung der Ein- und Ausgabegeräte inclusive der Erstellung von ICC-Geräteprofilen unter Normbedingungen und unter Verwendung von Farbmeßgeräten,
2. Integration der Profile in ein möglichst anwendungsprogrammunabhängiges CMS-Modul auf Betriebssystemebene,
3. cannen der Farbobjekte mit Eingabegeräten unter Verwendung des gerätespezifischen RGB-Farbraums in möglichst hoher Farbtiefe (24–48 Bit) unter Normbedingungen. Für die Direktdigitalisierung werden dabei Scanner für Aufsichtvorlagen, für die Filmdigitalisierung solche für Durchsichtvorlagen verwendet.[49] Wird vom Film digitalisiert, muß schon bei der Erstellung der Scanfilme auf Farbauthentizität geachtet werden durch eine Verfilmung unter Normbedingungen und unter Verwendung von Testvorlagen und Farbmeßgeräten.
4. Transformation der RGB-Daten in einen geräteunabhängigen Farbraum wie CIELAB unter Verwendung des CMS-Moduls und unter Verwendung eines plattformübergreifenden, nicht proprietären Dateiformats wie TIFF-unkomprimiert für die Speicherung als sogenannte digitale Master,
5. Transformation der digitalen Master in geräteabhängige Farbräume für die unterschiedlichen Anwendungbereiche und Ausgabeformen, z.B. nach RGB für die Bildschirmdarstellung im Internet[50] oder nach CMYK für den

---

[49] Aus Gründen der Bestandserhaltung sollte bei der Direktdigitalisierung mit vorlagenschonenden Buch- oder Kamerascannern und nicht mit Flachbettscannern gearbeitet werden. Die Digitalisierung vom Farbmikroformen (Rollfilm, Macrofiche) erfolgt momentan ebenfalls mit Kamerascannern mit einer speziellen Durchlichteinrichtung.

[50] Für Bilschirmdarstellungen könnte sich in Zukunft auch eine Transformation in den geräteunabängigen sRGB-Farbraum für sinnvoll erweisen. Zum sRGB-Farbraum siehe oben S. 186.

Druck unter gleichzeitiger Verwendung von anwendungsspezifischen Dateiformaten wie JPEG, PNG, TIF, EPS oder Postscript-Druckdateien.[51]

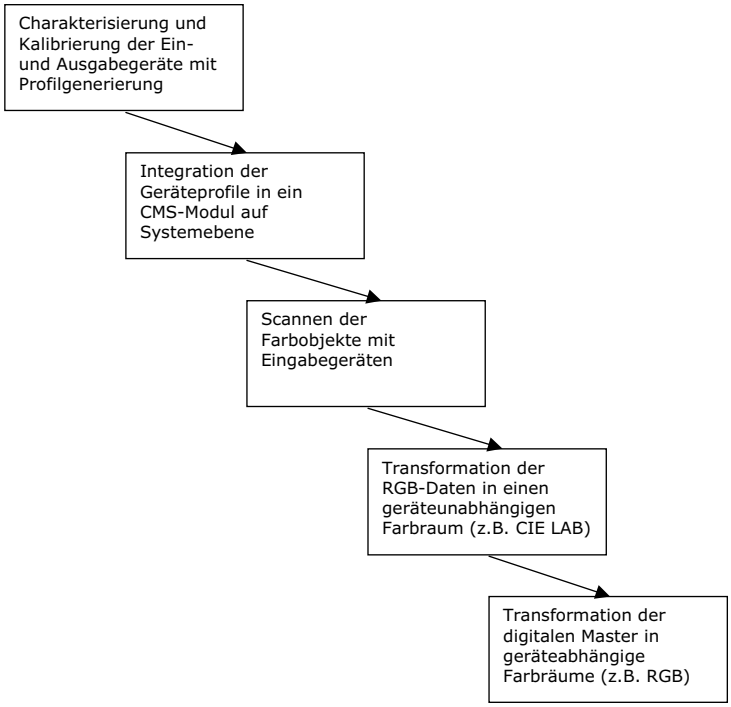


Abb. 6: Möglicher Workflow der Farbdigitalisierung unter Verwendung von Colormanagement.

Fazit und Ausblick

Zusammenfassend läßt sich feststellen, daß man im Rahmen eines *Colormanagements* mehrere verschiedene Faktoren berücksichtigen muß, um bei der Farbdigitalisierung erfolgreich zu sein. Neben der Beachtung der Umgebungsbedingungen, wie die Verwendung von Normlicht, charakterisierter

---

51 Dabei bieten sich für bildschirmorientierte Internetpräsentationen die Formate JPEG und PNG an, während für den Druck normalerweise die Dateiformate TIF und EPS verwendet werden.

und kalibrierter Geräte, Nutzung geräteunabhängiger Farbräume, geeigneter Dateiformate, Farbprofile und CMS-Software ist der Mensch als Anwender ein nicht zu unterschätzender Faktor innerhalb des CMS-Workflows. Damit der Ablauf reibungslos funktioniert, sind besondere Kenntnisse im Umgang mit Ein- und Ausgabegeräten sowie mit der verwendeten CMS-Software erforderlich.

Eine besondere Bedeutung kommt bei der Farbdigitalisierung von Archivgut bei einem *Outsourcing* dem Zusammenspiel zwischen dem Auftraggeber von Archivseite und dem Dienstleister zu. Hier müssen genaue Absprachen über die Verwendung von Farbräumen, Geräteprofilen und die CMS-Softwareumgebung getroffen werden.

Die Entwicklung und Anwendung von Colormanagement-Systemen befindet sich zur Zeit noch in der Aufbauphase. So gibt es für PC-Systeme auf Microsoft Windows-Basis bis auf die genormten ICC-Profile noch keine ausgereifte programmübergreifende und anwenderfreundliche Umgebung. Besser sieht es für die MacOS-Umgebung aus, wo mit ColorSync ein CMS-Systemmodul verankert ist. In der Praxis erfolgt allerdings der Farbdatenaustausch über heterogene Systemumgebungen hinweg, was momentan noch zu Problemen führt und bisher nur über die Nutzung von Anwendungsprogrammen wie Adobe Photoshop, das für beide Plattformen existiert, einigermaßen abgefangen werden kann.

Bestehende Colormanagement-Systeme werden in der Zukunft weiter verbessert werden, vor allem im Hinblick auf die Qualität der Farbraumtransformation, die Bedienerfreundlichkeit der Systeme, die Farbstabilität von Ein- und Ausgabegeräten und die Verbreitung von Standards und Anwendungssoftware, die diese Standards dann auch unterstützt.

Dennoch läßt sich schon heute sagen, daß sich durch den Einsatz von Colormanagement die Reproduzierbarkeit und Vorhersagbarkeit der Ergebnisse in jedem Arbeitsschritt einer Farbdigitalisierungs-Prozeßkette besser kontrollieren und nachvollziehen lassen als ohne. Farbmanagement ist somit ein wichtiger Beitrag zur Qualitätssicherung bei der Farbdigitalisierung von Archivgut.

Abbildungsnachweis

Abbildungen 1 und 2: Exakte Farbkommunikation. Vom Farbgefühl zur objektiven Messung. Broschüre der Minolta GmbH, Ahrensburg 1996. S. 10 und 13.

Abbildungen 3 und 4: Color Management. Die kreative Welt der digitalen Daten. Hg. von Linotype CPS. Bad Homburg o.J (PDF-Datei enthalten auf CD *Heidelberg. Color Publishing Solutions Version 4*), Art. Farbmessung.

Abbildungen 5 und 6: Gerald Maier.

Automatische Texterkennung bei digitalisiertem Archiv- und Bibliotheksgut

Von THOMAS FRICKE und GERALD MAIER

Durch verschiedene Testreihen im Rahmen des Forschungsprojekts *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Archivalien und Bücher* sollte gezeigt werden, ob und inwieweit eine automatische Texterkennung bei Archiv- und Bibliotheksgut möglich und sinnvoll ist. Die Behandlung von Archiv- und Bibliotheksgut mittels OCR-Verfahren[1] ist eine Grundvoraussetzung für Volltextindizierung bzw. Volltextrecherchemöglichkeiten auf der Grundlage digitaler Images.

Die Erprobung der automatischen Texterkennung erfolgte in Verbindung mit der Untersuchung der Qualität von 1 Bit und 8 Bit Digitalisaten. In verschiedenen Tests wurden das Leistungsvermögen und die Grenzen einer OCR-Texterkennung insbesondere bei der Digitalisierung von Mikrofilmen ermittelt. Hierzu wurden vom Projektteam OCR-Tests in Eigenleistung vorgenommen und verschiedene Tests als Dienstleistungen in Form eines *Outsourcings* vergeben. Untersucht wurden auch Lösungen zur Texterkennung von Frakturschriften. Zudem wurden Wirtschaftlichkeitsuntersuchungen angestellt, bei denen der automatischen Textumsetzung eine manuelle gegenübergestellt wurde.

Automatische Texterkennung in Abhängigkeit von der Bildqualität

Im Rahmen der Beschäftigung mit Fragen der Qualität von digitalen 1 Bit und 8 Bit-Images zeigte sich, daß die Bildqualität und die sie beeinflußenden Faktoren auch für die automatische Texterkennung von Bedeutung sind.[2]

Folgende Faktoren können Einfluß auf die Erkennungsrate haben:

- Scanauflösung
- Farbtiefe (1,4 oder 8 Bit) bzw. Filmvorlage (AHU, Halbton)
- Schwellenwert (bei 1 Bit)

---

1 OCR= Optical Character Recognition.

2 Siehe dazu in diesem Band den Beitrag von Gerald *Maier:* Qualität, Bearbeitung und Präsentation digitaler Bilder. S. 129.

- Qualität des Textbestandes
  - Schriftgröße
  - Konturschärfe
  - Buchstabendeformation (u.a. aufgebrochene Buchstaben, Ligaturen bzw. *verklebte* Buchstaben)
  - Einstreuung von handschriftlichen Vermerken oder graphischen Elementen
- erfolgtes bzw. nicht erfolgtes *Image Enhancement*
  - bei 8 Bit-Vorlagen zur Optimierung des Kontrastverhältnisses *Schrift – Vorlagenhintergrund*, (Tonwertkorrektur) und der Konturschärfe (Scharfzeichnen)
  - bei 1 Bit-Vorlagen zur Entfernung von Streupixeln
- Qualität der OCR-Software v.a. der Erkennungstechniken (mathematische Verfahren wie Omnifont-Technologie, linguistische Verfahren zur Rechtschreibprüfung unter Einsatz von Wörterbüchern, Trainierbarkeit)
- erfolgtes bzw. nicht erfolgtes Training der OCR-Software

Vom Projektteam wurden verschiedene Tests mit Objekten des *kleinen Warenkorbs* vorgenommen, bei denen zwei Aspekte untersucht wurden:

1. Vergleich der Erkennungsrate von Filmdigitalisaten und Direktdigitalisaten
2. Vergleich der Erkennungsraten von 1 Bit- und 8 Bit-Vorlagen

Verwendet wurde für die Tests die aus Rußland stammende OCR-Software FineReader 4.0 (Abbyy-Software), die sich durch folgende Fähigkeiten auszeichnet:

- hohe Erkennungsrate bei maschinenschriftlichen bzw. gedruckten Vorlagen
- Eignung für 1 Bit und 8 Bit Vorlagen
- nicht erkannte Bereiche werden als Grafik abgebildet
- Möglichkeit des Batch-Betriebs
- Trainierbarkeit

Test 1: Akten in Schreibmaschinenschrift

In einem ersten Test wurde die Erkennungsfähigkeit an maschinenschriftlichen Aktenschriftstücken erprobt.

Als Vorlage wurden digitale Formen eines Aktenschriftstücks aus dem *kleinen Warenkorb* verwendet.[3] Die Direktdigitalisierung wurde mit einem

[3] Objekt 03b: Bericht des Regierungspräsidiums Südbaden in Freiburg an das Innenministerium Baden-Württemberg in Stuttgart vom 8.6.1953, Rückseite.

3b

eingeführten Verfahren sehr gründlich durchgeführt werden, spricht, dass schon des öftern Löschfahrzeuge, die von Kraftfahrzeugsachverständigen überprüft und abgenommen worden waren, hier von den Feuerwehrfachleuten beanstandet werden mussten, weil die Fahrzeuge usw. feuerwehrtechnische Mängel aufgewiesen haben. Es wurde von den Beauftragten der Feuerwehrgeräte-Fabriken, die die Fahrzeuge zur Abnahme hierher brachten, auch schon wiederholt offen ausgesprochen, dass die Abnahmeprüfungen in Südbaden am schärfsten durchgeführt und dass Mängel, die hier beanstandet werden, anderwärts hingenommen werden würden; dabei wurde aber die Berechtigung der Beanstandungen zugegeben. ~~Gegen eine Abnahme der Geräte in den Herstellerwerken bestehen grundsätzliche Bedenken.~~ Abschliessend sei noch darauf hingewiesen, dass es nach den gemachten Erfahrungen ~~aus verschiedenen Gründen~~ besser ist, wenn die Geräte ~~nicht~~ im Herstellerwerk, sondern bei den Feuerwehrdienststellen abgenommen werden.

Das Regierungspräsidium würde es deshalb begrüssen, wenn die im Regierungsbezirk Südbaden eingeführte Regelung bestehen bliebe.

II. ZdA.

Abb. 1: Objekt 03b: Bericht des Regierungspräsidiums Südbaden in Freiburg an das Innenministerium Baden-Württemberg in Stuttgart vom 8.6.1953, Rückseite.

Flachbettscanner Microtek Scanmaker E6 vorgenommen, die Filmdigitalisierung mit einem Sunrise III-Filmscanner. Digitalisiert wurde jeweils im 1 Bit-Modus mit 600 dpi und im 8 Bit-Modus mit 400 dpi. Das Aktenschriftstück kann folgendermaßen charakterisiert werden:

- Schreibmaschinenschrift auf gedrucktem Briefkopf (gelbliches Papier mit handschriftlichen Vermerken in Bleistift, Kugelschreiber schwarzer und blauer Tinte)
- teilweise durchscheinende Schrift

Ein Vergleich der Direktdigitalisierung mit der Filmdigitalisierung bzw. ein Vergleich des Scannens im 1 Bit-Modus (600 dpi) und im 8 Bit-Modus (400 dpi) bringt folgendes Ergebnis nach manueller Fehlerauszählung:

Erkennungsrate (insgesamt 1143 Zeichen (1/2 Seite)):

- Direktscan 1 Bit: mehr als 31 Fehler: 97,3%
- Direktscan 8 Bit: 5 Fehler: 99,6%
- Filmscan 1 Bit: 7 Fehler: 99,4%
- Filmscan 8 Bit: 16 Fehler: 98,6%

Die Schreibmaschinenschrift (Courier, 12 Punkt) wird relativ fehlerlos übernommen, die handschriftlichen Vermerke dagegen werden nicht erkannt. Als problematisch erweist sich die Erkennung von Buchstaben mit unzureichendem Farbauftrag, bedingt durch unterschiedlichen Anschlag und Farbbandqualität.

Die teilweise durchscheinende Schrift führt zu einer erhöhten Fehlerrate v.a. im 1 Bit-Modus. Erstaunlicherweise ist beim Filmscan die Erkennungsrate bei 1 Bit höher als bei 8 Bit, während beim Direktscan ein genau umgekehrtes Verhältnis vorliegt. Die Erkennungsrate ist beim Filmscannen aber nur unwesentlich geringer.

Test 2: Amtsdrucksache

Bei der zweiten Testvorlage aus dem *kleinen Warenkorb* handelt es sich um ein gedrucktes Gesetzblatt mit Serifen-Schrift (Haupttext: 10 Punkt und Anmerkungen: 7 Punkt).[4] Die Direktdigitalisierung wurde ebenfalls mit einem Flachbettscanner Microtek Scanmaker E6 vorgenommen, die Filmdigitalisierung mit einem Sunrise III-Filmscanner. Wie bei Test 1 wurde jeweils im 1 Bit-Modus mit 600 dpi und im 8 Bit-Modus mit 400 dpi gescannt.

---

[4] Objekte 13a+b: Staats- und Verwaltungsrecht, Lieferung August 1971: Landesbeamtengesetz Baden-Württemberg C/1, S. 67f.

Abb. 2: Objekt 13a: Staats- und Verwaltungsrecht, Lieferung August 1971: Landesbeamtengesetz Baden-Württemberg C/1, S. 67.

Der Vergleich *Direktdigitalisierung – Filmdigitalisierung* und *1 Bit*–8 *Bit* bringt folgendes Ergebnis nach manueller Fehlerauszählung:

Erkennungsrate (insgesamt 4733 Zeichen (Vorder- und Rückseite)):

- Direktscan 1 Bit: 8 Fehler: 99,8%
- Direktscan 8 Bit: 9 Fehler: 99,8%
- Filmscan 1 Bit: 8 Fehler: 99,8%%
- Filmscan 8 Bit: 6 Fehler: 99,9%

In diesem Fall ist die Erkennungsrate nach der Fehlerzahl bei der filmdigitalisierten Version sogar minimal höher als bei der direktdigitalisierten, wobei prozentual eine gleiche Erkennungsrate vorliegt. Die filmdigitalisierte Version hat bei 8 Bit eine schlechtere Erkennungsrate als bei 1 Bit, was mit unausgeglichenen Tonwerten und unzureichender Konturschärfe der Buchstaben zusammenhängt.

OCR-Feldtest

Als Dienstleister für einen Feldtest wurde das Zentrum für Datenverarbeitung (ZDV) der Universität Tübingen ausgewählt, das eine entsprechende technische Infrastruktur und entsprechendes Know-how besitzt. Das ZDV verfügt über mehrere OCR-Programme, so daß deren unterschiedliche Leistungsfähigkeit und Eignung für entsprechende Vorlagen (u.a. Drucke mit verschiedenen Schriften, Schreibmaschine, Schreibmaschinen-Durchschlag) getestet werden konnte.

Als Testmaterial dienten 250 Images im TIF-/TIF-G4-Format. Dabei handelt es sich in erster Linie um eine Auswahl filmdigitalisierter Formen des *kleinen Warenkorbs*, die von verschiedenen Dienstleistern in unterschiedlichen Farbtiefen (1 Bit und 8 Bit), Auflösung (300, 400 und 600 dpi) und unterschiedlicher Scantechnik (Aufsichtscanner mit Durchlichteinheit und Filmscanner) erstellt wurden.[5] Ergänzt wurde das Testmaterial durch in Eigenleistung mit dem Flachbettscanner erstellte Direktscans der Objekte.

Folgende Objekte wurden aus dem *kleinen Warenkorb* ausgewählt:

01a Hiplers *12 Artikel*, S. 63 (Druck, Fraktur mit bitonaler Abbildung)
02a Ganzer, Führergesicht, S. 241 (Druck, Fraktur)

---

[5] Folgende Objekte aus dem *kleinen Warenkorb* wurden ausgewählt: 01a, 02a, 03a, 06a, 07a, 08, 11a, 13a, 16a und 17a.

03a Regierungspräsidium Südbaden in Freiburg an Innenministerium Baden-Württemberg in Stuttgart, 8.6.1953, Vorderseite (Schreiben, Schreibmaschine mit Vermerken in Bleistift und blauem Kugelschreiber)
06a Verwaltungsamt für Wirtschaft des amerikanischen und britischen Besatzungsgebiets in Minden: Anordnung über Preise für die Beförderung von Personen in Personenkraftwagen vom 18.10.1947, Vorderseite (Schreiben, Schreibmaschinen-Durchschlag in schwarz)
07a Finanzministerium Baden-Württemberg betr. Umstellungsgesetz, 27.07.1948, S. 1 (Schreiben, Schreibmaschine, Kopie)
08 Wilhelm Schoof, Annette von Droste-Hülshoff im Wandel der Zeiten, S. 8 (Druck, Antiqua)
11a Deutscher Bundestag, 3. Wahlperiode, Drucksache 1237, S. 71 (Druck, Antiqua)
13a Staats- und Verwaltungsrecht, Lieferung August 1971: Landesbeamtengesetz Baden-Württemberg C/1, S. 67 (Druck, Antiqua)
16a Jugend, Nr. 14, 1915, Vorderseite (Druck, verschiedene Schrifttypen und bitonale Abbildungen)
17a Ausschnitt aus Stuttgarter Zeitung Nr. 259 o.J., Vorderseite (Zeitungsdruck, verschiedene Schriften mit gerasterten Abbildungen)

Die Images wurden dabei in Testreihen organisiert, wobei jede Reihe alle o.g. Warenkorbobjekte beinhaltete und durch folgende Attribute näher bestimmt wurde:

1. Reihen-Nummer
2. Vorlagenart (Film, Originalobjekt)
3. Farbtiefe
4. Objektbezogene Scanauflösung
5. Dienstleister bzw. Scantechnik
6. Bearbeitungsstatuts (unbearbeitet/bearbeitet)

Um die Erkennungsrate verschiedener OCR-Software zu testen, wurden die ganzen Tests jeweils mit den aktuellen und ausgereiften Software-Produkten FineReader 4.0 (Abbyy) und Omnipage Professional 8.0 (Caere) durchgeführt. Zur Dokumentation und anschließenden Auswertung der Testergebnisse wurde ein spezielles Formular verwendet, in dem für jedes Image einer Testreihe folgende Angaben festzuhalten waren:

- Dateiname der mit OCR erzeugten Text-Datei,
- verwendete OCR-Software,
- Anzahl der Fehler,
- und Erkennungsrate in %.

Fehleranzahl und Erkennungsrate wurden im Anschluß an die OCR-Testläufe ermittelt und tabellarisch aufbereitet. Bei der Auswertung wurden die Objekte 01a, 02a, 06a, 16a nicht berücksichtigt, da hier schon optisch ersichtlich war, daß die Erkennungsrate aufgrund der Textbeschaffenheit der Vorlage (Fraktur, ineinanderlaufende Buchstaben) unter 80% lag. Bei den übrigen Objekten hat eine studentische Hilfskraft Fehleranzahl und prozentuale Erkennungsrate durch ein computerunterstütztes Verfahren für jedes Image ermittelt und im vorgegebenen Formular festgehalten. Gezählt wurden die richtig erkannten Buchstaben. Satzzeichen wurden als Buchstaben gewertet. Formatierungen wurden bei der Fehlererhebung nicht berücksichtigt.

Im Anschluß an die Vorarbeiten wurde der Feldtest unter folgenden Fragestellungen bzw. Aspekten ausgewertet:

1. Welche Objekte sind überhaupt geeignet für eine automatische Texterkennung?
2. Vergleich der Erkennungsrate bei 1 Bit-Vorlagen mit derjenigen bei 8 Bit-Vorlagen mit Einfluß der Farbtiefe, Scanauflösung und Filmvorlage (AHU, Halbton)
3. Vergleich der Erkennungsrate bei Filmdigitalisierung mit derjenigen bei der Direktdigitalisierung
4. Vergleich der Erkennungsrate bei unbearbeiteten Images mit derjenigen nachbearbeiteter Images (Schärfen, Tonwertkorrektur)
5. Vergleich der Erkennungsrate unterschiedlicher Standard-OCR-Software (FineReader 4.0 und Omnipage Professional 8.0)

Zu 1: Für eine automatische Texterkennung mit Standardsoftware sind grundsätzlich nur Objekte mit Druckschriften geeignet wie maschinenschriftliche Akten (03a, 07a), Buchdruck (08), Drucksachen (11a, 13a) oder Zeitungen (17a). Für maschinenschriftliche Akten eignet sich eine OCR-Erfassung nur bedingt, da hier meist handschriftliche Vermerke eingestreut sind (03a), die eine Erkennung auch des maschinenschriftlichen Umgebungstextes oft unmöglich machen. Ungeeignet ist OCR auch bei der Erfassung von Schreibmaschinendurchschlägen (06a), da hier die Buchstaben oft deformiert, verklebt (ineinanderfließende Konturen) oder aufgebrochen sind.

Handschriften, auch in Form von Vermerken (03a), und Fraktur (01a, 02a) können ohne Spezialtraining beim momentanen technischen Stand durch Standard-OCR-Software nicht erkannt werden. Zwar bietet heutige Standard-Software wie FineReader 4.0 in der Handprint-Version auch eine Texterkennung von Handschriften an. Dabei können allerdings nur gedruckte Großbuchstaben in Formularen erkannt werden, nicht aber Kleinbuchstaben und Schreibschriften.

Problematisch sind auch Dokumente mit komplizierter Seitenstruktur und Abbildungen, die von Texten umflossen werden, wie sie z.B. bei Anzeigentei-

len von Zeitungen auftreten (16a). Erschwerend für eine Erkennung wirken sich auch der Wechsel verschiedener Schrifttypen (16a), ungleichmäßiger Druck mit unterschiedlichem Kerning, schrägen Zeilen oder unterschiedlichem Durchschuß (08) in einem Dokument aus.

Zu 2: Grundsätzlich führte eine Erhöhung der Eingabeauflösung von 300 dpi auf 400 dpi zu einer besseren Erkennungsrate. Eine Erhöhung der Auflösung von 400 dpi auf 600 dpi bringt dagegen eine nur geringfügige Verbesserung der Erkennungsrate, teilweise führte sie sogar zu einer Verschlechterung (siehe Reihe 12, Objekt 17a). Bei der Filmdigitalisierung mit Sunrise-Filmscannern konnten mit 8 Bit Images, gescannt vom Halbtonfilm, bei vergleichbarer Auflösung bessere Ergebnisse erreicht werden als mit 1 Bit Images, gescannt vom AHU-Film. Dies gilt allerdings nur für eine Scanauflösung von bis zu 400 dpi. Ab 600 dpi werden bei 8 Bit wieder schlechtere Ergebnisse erzielt als bei 1 Bit Farbtiefe mit 400 und 600 dpi. Vergleichbare Erkennungsraten werden mit 600 dpi/1 Bit und 400 dpi/8 Bit erreicht, wobei 1 Bit-Images aufgrund ihrer wesentlich geringeren Dateigröße vorzuziehen sind.[6]

Reihe	Film	Bit	dpi	Scantechnik	7a	8	11	13a	17a	Mittelwert
9	negativ	1	300	Sunrise ProScan	86,18%	94,17%	99,10%	97,85%	98,89%	95,24%
11	negativ	1	400	Sunrise ProScan	91,24%	98,66%	99,55%	98,33%	99,16%	97,39%
12	negativ	1	600	Sunrise ProScan	92,66%	98,75%	99,61%	98,33%	98,18%	**97,51%**
16	positiv	8	400	Sunrise ProScan	92,87%	98,93%	99,61%	97,01%	99,03%	**97,49%**
17	positiv	8	600	Sunrise ProScan	90,91%	98,89%	99,27%	95,35%	98,15%	96,51%

Abb. 3: Vergleich der Erkennungsrate bei der Filmdigitalisierung mit Sunrise Filmscannern bei 1 Bit und 8 Bit-Digitalisaten mit FineReader 4.0.

Zu 3: Der Feldtest bestätigte im Grunde das Ergebnis der schon zuvor in Eigenleistung vom Projektteam durchgeführten OCR-Tests. Die Erkennungsrate bei der Filmdigitalisierung liegt nur unwesentlich unterhalb derer bei der Direktdigitalisierung mit Aufsichtscantechnik, teilweise werden mit der Filmdigitalisierung sogar bessere Resultate erzielt.

---

6 8 Bit Graustufen-Bilddateien in 400 dpi sind etwa 3,5 mal so groß als 1 Bit Bilder in 600 dpi Auflösung.

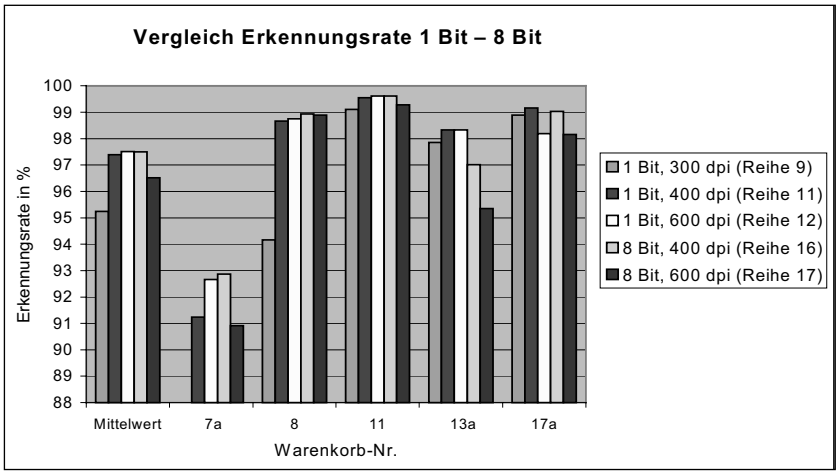


Abb. 4: Vergleich der Erkennungsrate bei der Filmdigitalisierung mit FineReader 4.0.

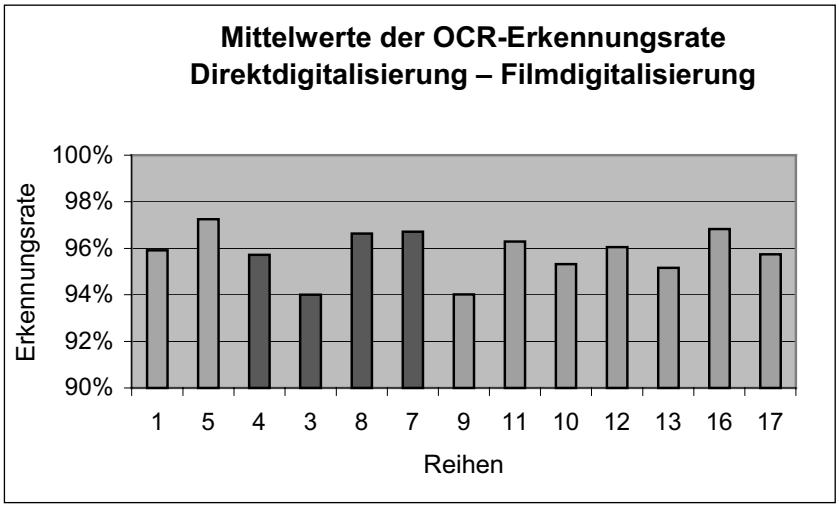


Abb. 5: Mittelwerte der Erkennungsraten bei verschiedenen Scantechnologien.

Reihe	Direkt	Bit	dpi	Scantechnik	3a	7a	8	11a	13a	17a	Mittelwert
1		1	600	Aufsichtscanner mit CCD-Kopf, 8 Bit in 1 Bit gerechnet	91,59%	89,23%	98,22%	99,18%	98,43%	98,82%	95,91%
5		8	400	Aufsichtscanner mit CCD-Kopf	93,21%	93,20%	98,70%	99,63%	99,80%	98,97%	97,25%
4		1	300	Microtek Scanmaker E6 Flachbettscanner	93,45%	86,94%	98,24%	99,69%	97,89%	98,13%	95,72%
3		1	400	Microtek Scanmaker E6 Flachbettscanner	93,91%	76,44%	98,35%	99,47%	97,94%	97,92%	94,01%
8		8	300	Microtek Scanmaker E6 Flachbettscanner	93,74%	92,00%	98,96%	99,66%	98,19%	97,26%	96,64%
7		8	400	Microtek Scanmaker E6 Flachbettscanner	93,74%	92,27%	98,96%	99,58%	98,38%	97,36%	96,72%
Reihe	Film	Bit	dpi	Scantechnik	3a	7a	8	11a	13a	17a	Mittelwert
9	neg.	1	300	Sunrise ProScan III	87,88%	86,18%	94,17%	99,10%	97,85%	98,89%	94,01%
11	neg.	1	400	Sunrise ProScan IV	90,78%	91,24%	98,66%	99,55%	98,33%	99,16%	96,29%
10	neg.	1	600	Sunrise ProScan III	90,60%	89,93%	99,14%	99,47%	97,94%	94,90%	95,33%
12	neg.	1	600	Sunrise ProScan IV	88,75%	92,66%	98,75%	99,61%	98,33%	98,18%	96,05%
13	pos.	8	400	Sunrise ProScan III	87,94%	91,95%	94,58%	99,49%	98,14%	98,88%	95,16%
16	pos.	8	400	Sunrise ProScan IV	93,56%	92,87%	98,93%	99,61%	97,01%	99,03%	96,84%
17	pos.	8	600	Sunrise ProScan IV	91,88%	90,91%	98,89%	99,27%	95,35%	98,15%	95,74%

Abb. 6: Vergleich der Erkennungsraten bei verschiedenen Scantechnologien (Aufsicht-, Flachbett- und Filmscanner) mit FineReader 4.0.

Zu 4: Durch das Projektteam wurde auch untersucht, ob die OCR-Erkennungsrate durch eine Bildnachbearbeitung vor dem OCR-Durchgang verbessert werden kann.[7] Das Ergebnis des Feldtests zeigt, daß durch eine Nachbearbeitung von 1 Bit Images mit der Bildbearbeitungssoftware PixEdit keine Verbesserung der Ergebnisse erreicht werden kann. Der dort eingesetzte Filter wirkt global auf das gesamte Image und entfernt nicht nur störende Informationen wie Streupixel, sondern auch Bestandteile der Buchstaben (z.B. Serifen, Punkte), so daß selbst bei behutsamer Anwendung des Filters eher eine Verschlechterung der OCR-Ergebnisse eintritt, obwohl der Einsatz dieses Filters auf den ersten Blick eine scheinbare Verbesserung der Bildqualität bringt.

Auch bei der Nachbearbeitung von 8 Bit Images durch Tonwert- und Schärfenkorrektur konnte keine Verbesserung der OCR-Erkennungsrate erreicht werden. Auch hier wirken sich die Verbesserungen nur visuell aus (vgl. Reihen 22 mit Reihen 16).

unbearbeitet:

Reihe	Film	Bit	dpi	Scantechnik	3a	7a	8	11a	13a	17a
9	negativ	1	300	Sunrise ProScan III	87,88%	86,18%	94,17%	99,10%	97,85%	98,89%
10	negativ	1	600	Sunrise ProScan III	90,60%	89,93%	99,14%	99,47%	97,94%	94,90%
13	positiv	8	400	Sunrise ProScan III	87,94%	91,95%	94,58%	99,49%	98,14%	98,88%

bearbeitet:

Reihe	Film	Bit	dpi	Scantechnik	3a	7a	8	11a	13a	17a
18a	negativ	1	300	Sunrise ProScan III (nachbearbeitet mit PixEdit 4.0, Filter *Linien von bis*: 1–2)		89,72%	97,43%	99,21%		98,65%
18b	negativ	1	300	Sunrise ProScan III (nachbearbeitet mit PixEdit 4.0, Filter *Linien von bis*: 1–3)		89,77%	99,03%	99,04%		98,63%
19	negativ	1	600	Sunrise ProScan III (nachbearbeitet mit PixEdit 4.0, Filter *Linien von bis*: 1)		89,17%	99,00%			
22	positiv	8	400	Sunrise ProScan III (Tonwertkorrektur und Scharfzeichnen m. Thumbs Plus 3.2)	90,08%	89,45%	93,15%	98,31%	98,14%	77,35%

[7] Zu den Methoden des *Image Enhancements* und den dazu verwendeten Software-Werkzeugen siehe in diesem Band den Beitrag von Gerald *Maier:* Qualität, Bearbeitung und Präsentation digitaler Bilder. S. 129.

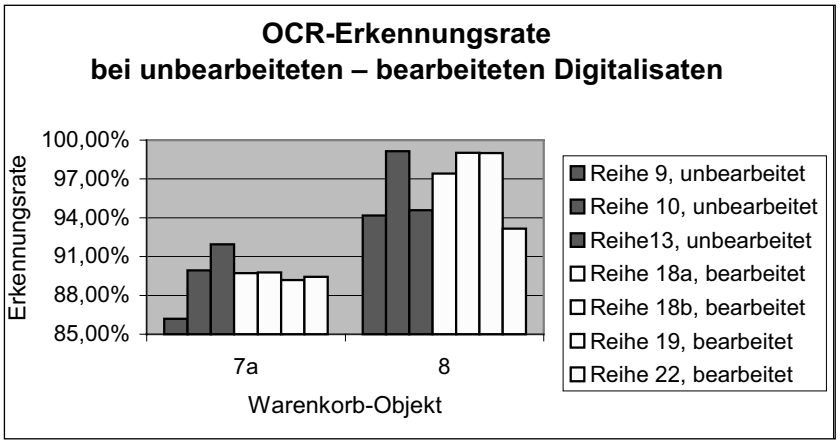


Abb. 7: Vergleich der OCR-Erkennungsrate bei unbearbeiteten und nachbearbeiteten filmdigitalisierten Images.

Zu 5: Der Test hat auch gezeigt, daß je nach eingesetzter OCR-Software unterschiedliche Erkennungsraten erreicht werden. Der Vergleich zwischen den Programmen FineReader 4.0 und Omnipage Professional 8.0 zeigt, daß FineReader insgesamt eine höhere Erkennungsrate aufweist.

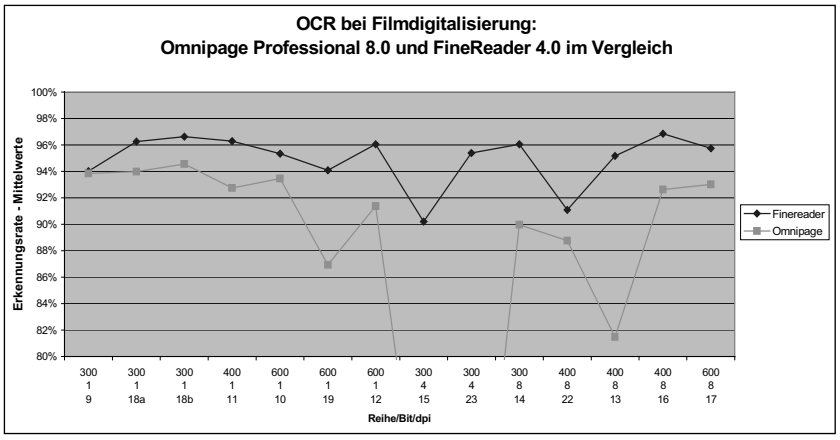


Abb. 8: OCR-Feldtest mit Objekten des kleinen Warenkorbs: Vergleich der Erkennungsrate von OCR-Software: FineReader 40 – Omnipage Professional 8.0.

OCR-Tests mit Frakturschriften

Durch das Projektteam wurde ein Test der OCR-Erkennung von Frakturschriften am Beispiel von Theaterzetteln aus den Beständen der Württembergischen Landesbibliothek[8] mit der Software FineReader 4.0 Handprint durchgeführt. Dabei wurde das Texterkennungsprogramm mit vier Seiten à 1 500 Zeichen trainiert. Der Gesamtzeitaufwand für das Training betrug drei Stunden.

Bei der anschließenden OCR-Erfassung wurde eine Erkennungsrate von 82,6% ermittelt, was für eine sinnvolle automatische Texterkennung nicht akzeptabel ist. Bessere Ergebnisse könnten sicherlich erreicht werden, wenn das Training auf einen größeren Textbestand ausgedehnt werden würde. Dabei muß in Betracht gezogen werden, daß jede Frakturschrift-Type neu zu trainieren ist.

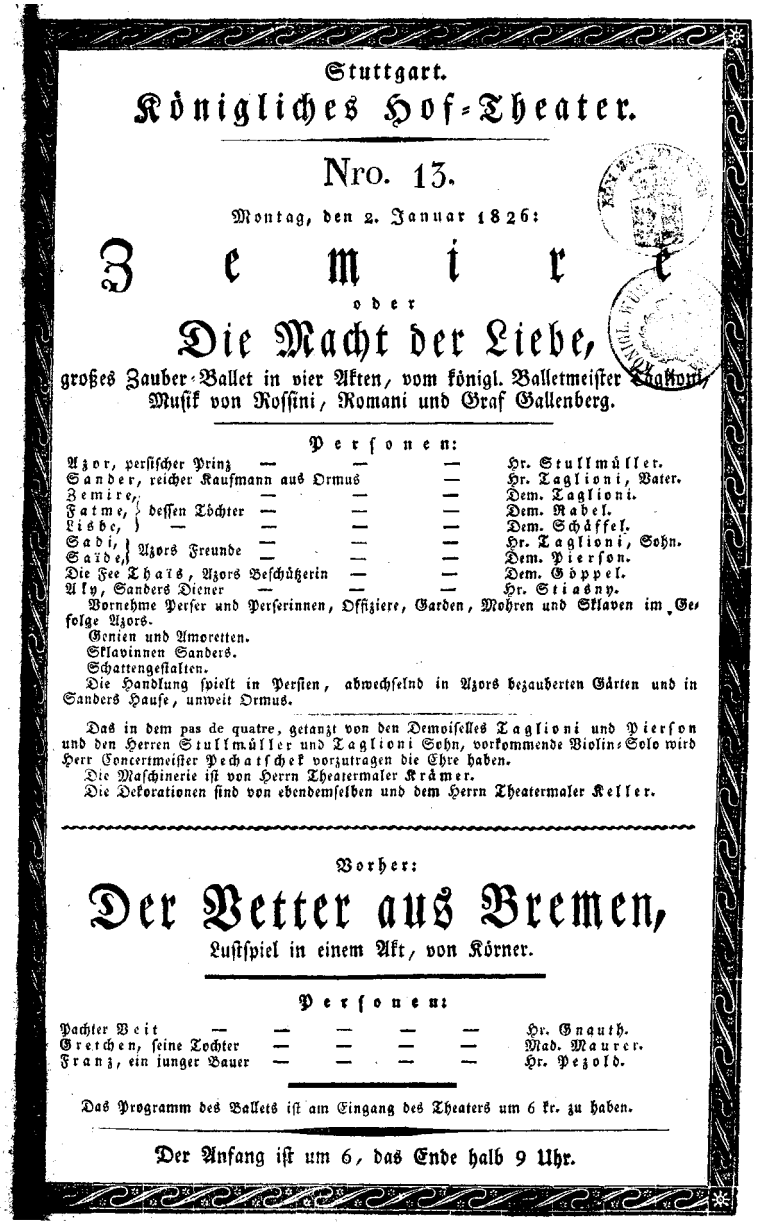
Folgende Erkennungsprobleme sind aufgetreten, die durch manuelles Training auch nur bedingt behoben werden können:

- Die Unterscheidung von langem *s* und *ſ* ist bei Frakturschrift generell schwierig, da sich beide Buchstaben beim Lesen oft nur durch den Wortsinn unterscheiden lassen und hier immer ein Fehlerrisiko zu erwarten ist. Ligaturen sind für Finereader dagegen kein Problem.
- Initialen sind oft verziert und werden gar nicht als Buchstaben wahrgenommen.
- Oft liegt aufgrund der manuellen Satz- und Drucktechnik ein unsauberes Schriftbild vor (unklare Konturen, ungleichmäßige Verteilung der Druckerschwärze, aufgebrochene Buchstaben, schlechtes Kontrastverhältnis).
- Häufig hat die Vorlage ein uneinheitliches und kompliziertes Layout mit wechselnder Satz- und Zeilenstruktur (häufiger Schrifttypen- und Schriftgrößenwechsel). Sehr große Überschriften werden manchmal nicht als Einheit erkannt, Zeilen sind vertikal oft nur geringfügig gegeneinander versetzt.
- Druckbild, Schrifttypen und Layout ändern sich von Vorlage zu Vorlage.
- Die vielen verschiedenen Frakturtypen machen einen unvertretbar hohen Trainieraufwand erforderlich.

Bessere Ergebnisse bei der Texterkennnung von Frakturschriften, als mit Standardsoftware, lassen sich durch die Verwendung von speziell für solche Schrifttypen entwickelter Software oder durch Anpassung von Standardsoftware erzielen.

---

[8] *Großer Warenkorb* LK314 001-005: *Theaterzettel: Württemberg. Staatstheater 1826* (Württembergische Landesbibliothek, Za 4914–1974/75; Musik-LS).



Stuttgart.

Königliches Hof-Theater.

Nro. 13.

Montag, den 2. Januar 1826:

Zemire

oder

Die Macht der Liebe,

großes Zauber-Ballet in vier Akten, vom königl. Balletmeister Taglioni,
Musik von Rossini, Romani und Graf Gallenberg.

Personen:

Azor, persischer Prinz	—	—	—	Hr. Stullmüller.
Sander, reicher Kaufmann aus Ormus			—	Hr. Taglioni, Vater.
Zemire,	—	—	—	Dem. Taglioni.
Fatme, } dessen Töchter	—	—	—	Dem. Rabel.
Lisbe, } —	—	—	—	Dem. Schäffel.
Sadi, } Azors Freunde	—	—	—	Hr. Taglioni, Sohn.
Saïde, }	—	—	—	Dem. Pierson.
Die Fee Thaïs, Azors Beschützerin		—	—	Dem. Göppel.
Aly, Sanders Diener	—	—	—	Hr. Stiasny.

Vornehme Perser und Perserinnen, Offiziere, Garden, Mohren und Sklaven im Gefolge Azors.
Genien und Amoretten.
Sklavinnen Sanders.
Schattengestalten.
Die Handlung spielt in Persien, abwechselnd in Azors bezauberten Gärten und in Sanders Hause, unweit Ormus.

Das in dem pas de quatre, getanzt von den Demoiselles Taglioni und Pierson und den Herren Stullmüller und Taglioni Sohn, vorkommende Violin-Solo wird Herr Concertmeister Pechatschek vorzutragen die Ehre haben.
Die Maschinerie ist von Herrn Theatermaler Krämer.
Die Dekorationen sind von ebendemselben und dem Herrn Theatermaler Keller.

Vorher:

Der Vetter aus Bremen,

Lustspiel in einem Akt, von Körner.

Personen:

Pachter Veit	—	—	—	—	Hr. Gnauth.
Gretchen, seine Tochter		—	—	—	Mad. Maurer.
Franz, ein junger Bauer		—	—	—	Hr. Pezold.

Das Programm des Ballets ist am Eingang des Theaters um 6 kr. zu haben.

Der Anfang ist um 6, das Ende halb 9 Uhr.

Abb. 9: *Theaterzettel: Württemberg. Staatstheater 1826* (Württembergische Landesbibliothek, Za 4914–1974/75; Musik-LS).

Solche Spezialsoftware auf Basis des freien UNIX-Betriebssystems Linux wird z.B. von einer Potsdamer Firma entwickelt. Auch hier muß jede Fraktur-Schrifttype einzeln in einer digitalen Vorverarbeitungsstufe trainiert werden. Erfolge werden mit dieser Software besonders bei verklebten und zusammenhängenden Buchstaben erzielt, die normalerweise eine Texterkennung unmöglich machen, da diese auf eine klare Trennung der Zeichen angewiesen ist. Durch mathematische Berechnungen können hier zusammenhängende Buchstaben auf Softwareebene getrennt werden. Der Einsatz solcher Spezialsoftware erfordert aber eine ständige individuelle Programmanpassung und lohnt sich erst bei einem größeren Auftragsvolumen. Die Firma verspricht eine Erkennungsrate von bis zu 99,8%. Derzeit beschränkt sich der Einsatzbereich dieser Entwicklung noch auf die Dienstleistung nach folgendem Ablauf:

1. Anpassung und Training der OCR-Software anhand der ersten 10 Seiten
2. OCR-Erfassung aller Seiten
3. Manuelle Nachkorrektur (falls notwendig)

Eine Anpassung von Standardsoftware für die Erkennung von Frakturschriften wird z.B. von einem Berliner Dienstleister vorgenommen. Hier wird die Standardsoftware FineReader durch ein Zusatzprogramm erweitert, das den Nachbearbeitungsaufwand verringert. Die Texte bleiben weiterhin mit FineReader editierbar. Das Verfahren funktioniert nach Auskunft des Anbieters bei Filmdigitalisaten allerdings nur bei guten Kontrastverhältnissen und nicht deformierten Lettern.

Um die Leistungsfähigkeit des Verfahrens zu testen, wurde ein Test durch den Dienstleister mit den zwei folgenden filmdigitalisierten Objekten durchgeführt:

1. Als erstes Testobjekt diente ein Buchdruck in Fraktur mit guten Kontrastverhältnissen aus dem *kleinen Warenkorb* (Objekt 01a). Die 1 Bit-TIFF-Datei, gescannt mit 600 dpi, wurde mit der OCR-Software FineReader 3.0b bearbeitet. Zunächst wurde der zu erkennende Textblock für die Erkennung markiert, dann wurde ein Zeichentraining vorgenommen. Anschließend wurde mit der Datei ein erneuter OCR-Erkennungsdurchgang vorgenommen. Danach wurde mittels eines selbstentwickelten Filters eine automatisierte Nachbearbeitung durchgeführt, die typischen Fraktur-OCR-Fehler und Finereader-Bugs beseitigt.

Der Text wurde fast fehlerlos erkannt. Allerdings ist aufgrund der sehr beschränkten Datenmenge für den Test das Ergebnis wenig aussagekräftig. Für ein sinnvolles Training ist ein Bestand von 10 000 Zeichen Voraussetzung. Nach Aussage des Dienstleisters sei aber generell bei einer Textvorlage mit dieser Druckqualität mit dem eingesetzten Verfahren eine sehr geringe Fehlerquote (Erkennungsrate bis zu 99,9%) zu erreichen. Ein Korrekturlesen und

eine eventuelle manuelle Nachkorrektur ist normalerweise zusätzlich erforderlich.

2. Weniger erfolgreich war der Test mit den Theaterzetteln aus den Beständen der Württembergischen Landesbibliothek. Auch hier zeigte sich, wie schon bei dem durch das Projektteam durchgeführten Test, daß diese Vorlagen aus den schon genannten Gründen für OCR-Erfassung mit heutiger Standardsoftware nicht geeignet sind.

Wirtschaftlichkeitsuntersuchung *Texterfassung mittels OCR – händige Texterfassung*

Grundlage für die Wirtschaftlichkeitsuntersuchung sind der Feldtest und eine Marktanalyse.

Folgende Daten müssen im Rahmen einer Wirtschaftlichkeitsanalyse erhoben werden, um Vergleiche zwischen verschiedenen Verfahren und Angeboten der automatischen Texterkennung mit denjenigen einer händigen Texterfassung anstellen zu können:

- Zeitaufwand für einen OCR-Durchgang pro Image, pro DIN A4-Seite oder pro 1000 Zeichen
- Erkennungsrate in % pro Image
- Zeitaufwand für Korrekturen eines OCR-Durchgangs pro Image, pro DIN A4-Seite oder pro 1000 Zeichen
- Zeitaufwand für sonstige Arbeiten bzw. Arbeitsschritte pro Image (v.a. Einrichten oder Trainieren der OCR-Software)
- Zeitaufwand für komplette händige Texterfassung pro Image, DIN A4-Seite oder pro 1000 Zeichen
- Kosten für OCR-Operator pro Stunde, pro Zeichen oder pro Image
- Kosten für Schreib-/Lektoratskraft pro Stunde, pro Zeichen oder pro Image
- Gemeinkosten für die Nutzung einer technischen Infrastruktur wie Räumlichkeiten oder technischer Ausstattung

Wirtschaftlichkeitsanalyse anhand des Feldtests

Um Anhaltspunkte für die Wirtschaftlichkeit von OCR-Verfahren zu bekommen, wurde ein Teil der OCR-Feldtestergebnisse einem weiteren Test unterzogen.

Der automatischen Textumsetzung wurde eine manuelle gegenübergestellt. Eine Schreibfachkraft hat an ausgewählten Beispielen zunächst die OCR-Ergebnisse in der entsprechenden Textdatei nach der Originalvorlage korrigiert. Anschließend hat sie den Text komplett in eine neue Datei abgeschrieben. Bei jedem Vorgang wurde die benötigte Zeit festgehalten.

Folgende Parameter wurden für die Berechnung zugrunde gelegt:

- Zeichen pro DIN A4 Seite: ca. 1000 Zeichen
- Kosten Schreib-/Lektoratskraft: DM 21,00/Stunde = DM 0,35/Min
- Kosten OCR-Operator: DM 21,00/Stunde = DM 0,35/Min

Bei der Wirtschaftlichkeitsberechnung wurden der Einfachheit halber die sog. Gemeinkosten außer Betracht gelassen, die aber durchaus für die Entscheidung *Outsourcing* oder Ausführung der automatischen Texterkennung in Eigenleistung von Bedeutung sein können. Daher können die hier ermittelten Ergebnisse nicht einfach mit Angeboten von Dienstleistern verglichen werden.

Die Ermittlung des OCR-Zeitbedarfs pro Testimage basiert auf dem Einsatz eines Standard-PCs mit Pentium II–400Mhz-Prozessor, 128 MB Hauptspeicher, und der Verwendung von FineReader 4.0 Standard als OCR-Software.

Nr.	Testreihe und Objekt	Erkennungsrate OCR in %	Zeitaufwand OCR-Erfassung	Zeitaufwand OCR-Korrektur	Zeitaufwand händige Texterfassung	Kosten für OCR-Erfassung incl. OCR-Korrektur in DM	Kosten für händige Texterfassung in DM
1	Reihe 16, 03a	87,53	2 Min	18 Min	9 Min	7,00	3,15
2	Reihe 12, 08	93,54	1 Min 30 Sec	22 Min	29 Min	8,23	10,15
3	Reihe 12, 13a	99,41	1 Min 30 Sec	8 Min	19 Min	3,33	6,65

Abb. 10: Vergleich OCR-Erfassung – händige Texterfassung.

Der Test hat gezeigt, daß bei einer Erkennungsrate unter 90% ein sehr hoher Nachbearbeitungsaufwand für Korrekturen besteht, so daß auf jeden Fall eine händige Texterfassung wirtschaftlicher ist als eine automatische. So sind die Kosten für eine automatische Texterkennung und anschließender manueller Korrektur bei Testobjekt Nr. 1 mit einer Erkennungsrate von 87,53% mehr als doppelt so hoch als bei einer manuellen Erfassung. Umgekehrt sieht es bei Testobjekt Nr. 3 aus, wo die Kosten für die automatische Texterkennung bei einer erzielten Erkennungsrate von 99,41% genau halb so hoch sind wie bei einer händigen Texterfassung. Schwieriger verhält es sich bei Erkennungsraten um 95%. Hier liegen die Kosten für eine OCR-Erfassung und händige Texterfassung nahe beieinander.

Die Frage, ob eine automatische Texterkennung oder eine händige Textumsetzung wirtschaftlicher ist, ist v.a. von den beiden folgenden Faktoren abhängig:

1. von der Beschaffenheit der Vorlage und der damit zu erzielenden Erkennungsrate,
2. von der Größe des Textvolumens, die bei einer Auftragsvergabe an Dienstleister die Preisgestaltung maßgeblich mitbestimmt.

Unabhängig von der vorgenommenen Wirtschaftlichkeitsanalyse sind die aktuelle Marktlage und die tatsächlichen Angebote verschiedener Dienstleister für die automatische Texterkennung und die händige Textumsetzung miteinander zu vergleichen, da für eine Kostenkalkulation oft das zu bearbeitende Textvolumen entscheidend ist, was in diesem Test nicht berücksichtigt werden konnte. Nicht berücksichtigt werden konnte auch der Einsatz zusätzlicher linguistischer automatischer Korrekturhilfen bei der OCR-Erkennung, die über die eingebauten Funktionen der Texterkennungssoftware hinausgehen. Dadurch könnte eine Verbesserung der Erkennungsrate erzielt werden und somit die Kosten für eine manuelle Nachkorrektur vermindert werden.

Wirtschaftlichkeitsanalyse aufgrund der tatsächlichen Preisgestaltung bei Dienstleistern

Um die tatsächliche Preisgestaltung bei Dienstleistern unabhängig von den eigenen Tests zu ermitteln, wurden verschiedene Angebote sowohl für die automatische Texterfassung als auch die manuelle eingeholt.

Die Preisgestaltung für eine OCR-Erfassung wird von einzelnen Dienstleistern teilweise unterschiedlich gehandhabt. Oft wird zunächst eine Grundgebühr berechnet, die ggf. das Training eines Schriftfonts beinhaltet, dann wird zumeist nach dem Zeichenvolumen (1000 Zeichen/Seite) abgerechnet. Die Kosten pro 1000 Zeichen sind normalerweise abhängig vom Gesamtauftragsvolumen. Je höher dieses ist, je günstiger ist der Preis für 1000 Zeichen bzw. für eine Seite. Kosten für eine Nachkorrektur werden normalerweise gesondert berechnet.

Folgende Angebote für die OCR-Texterfassung wurden bei verschiedenen Dienstleistern für normal gedruckte Vorlagen ohne handschriftliche Vermerke und Frakturtype eingeholt:

1. DM 0,80/1000 Zeichen bei einem Textvolumen von ca. 500 000 Seiten (Stand: Juni 1999)
2. DM 2,00–3,00/1000 Zeichen, d.h. ca. DM 2,00–3,00/DIN A4 Seite (Stand: Mai 1998)
3. DM 0,65–9,50/1000 Zeichen (ohne Korrektur und interaktives Training) (Stand: August 1998)

Dabei bewegt sich die Preisspanne bei den Angeboten zwischen DM 0,65–9,50/1000 Zeichen, wobei schon für unter DM 1,00/1000 Zeichen eine automatische Texterkennung angeboten wird.

Zusätzlich wurden speziell für OCR mit Frakturschriften bei zwei Dienstleistern besondere Angebote eingeholt. Auch hier wird unterschieden zwischen Grundpreis und Zusatzkosten für das Training weiterer Schrifttypen:

1. DM 5,80/1000 Zeichen (incl. Grundkosten und Training von zwei Fonts bei einem Gesamtvolumen von 750 Seiten à 1000 Zeichen); DM 1200,00 Anpassung der OCR auf jeden weiteren Schriftfont (Stand: Mai 1999) => DM 5,80/1000 Zeichen
2. DM 1,20/1000 Zeichen (incl. Grundkosten und Training für einen Schriftfont bei einem Gesamtvolumen von 500 Seiten à 1000 Zeichen) (Stand: Juli 1999) => DM 1,20/1000 Zeichen

Die Kosten für die Erkennung von Frakturschriften (maximal zwei verschiedene Typen) liegen nach diesen Angeboten ohne Nachkorrektur zwischen DM 1,20–5,80/1000 Zeichen bei einer zugesagten Erkennungsrate von 99,8–99,9%. Die deutlichen Preisunterschiede bei den Angeboten ergeben sich aus der Verschiedenheit und Aufwendigkeit der eingesetzten Erkennungsverfahren. So ist der Dienstleister des ersten Angebots aufgrund einer vollständig neu entwickelten Spezialsoftware auch in der Lage, vom Layout kompliziertere und inhomogenere Vorlagen mit unterschiedlichen Frakturschrifttypen zu bearbeiten, was sich in einem höheren Preis niederschlägt. Auch hier sind Kosten pro 1000 Zeichen normalerweise vom Gesamtvolumen abhängig, so daß bei einem größeren Volumen die Kosten pro Seite bzw. 1000 Zeichen geringer ausfallen.

Im Rahmen der Wirtschaftlichkeitsuntersuchung wurde auch ein Angebot zur händigen Texterfassung eingeholt. So bietet ein chinesischer Dienstleister ein manuelles Erfassen von Papiervorlagen an. Dabei werden alle Texte manuell zweimal je von unterschiedlichen Personen abgeschrieben und anschließend computerunterstützt auf Diskrepanzen hin analysiert. Der Dienstleister garantiert dabei eine Fehlerrate von unter 2/10 000 (Erkennungsrate von 99,98%).

Der Preis für diese Art der Texterfassung beträgt je nach Vorlage und Textvolumen zwischen ca. DM 1,30–1,50/1000 Zeichen.

Aufgrund der Marktanalyse und der ermittelten Preise kann keine generelle Empfehlung für eines der genannten Verfahren gegeben werden. Da die manuelle Texterfassung zu diesen Preiskonditionen nur im asiatischen Raum angeboten wird, kommen zusätzliche höhere Transportkosten hinzu, die eventuell durch eine elektronische Datenübertragung gemindert werden könnten. Das Transportrisiko, das bei einer Übersendung von Archivalien und Bibliotheksgut bestehen würde, entfällt bei der Verwendung von Filmdigitalisaten.

Mit Handschriften ist eine automatische Texterkennung bisher nicht durchführbar. Hier ist die manuelle Textumsetzung die einzige Möglichkeit, den Text in maschinenlesbarer Form zu erhalten. Allerdings dürften bei handschriftlichen Vorlagen oder bei Vorlagen mit älteren Frakturschriften und kompliziertem Seitenlayout auch Dienstleister einer manuellen Textumsetzung überfordert sein, da diese Schriften eine besondere Lesequalifikation der Schreibkräfte des Dienstleisters erfordern, die nicht unbedingt vorausgesetzt werden kann.

Schluß

Insgesamt kann festgehalten werden, daß der OCR-Einsatz bisher nur bei gedrucktem und maschinenschriftlichem Archiv- und Bibliotheksgut möglich ist. Handschriften werden aufgrund ihrer Individualität und der verwendeten Schreibschriften beim bisherigen Stand der Technik nicht brauchbar erkannt. Zwar bieten heute selbst Standardsoftware-Produkte wie FineReader 4.0 Handprint eine Erkennung von handschriftlichen Buchstaben an. Diese beschränkt sich aber auf sauber geschriebene Druckbuchstaben in Form von Großbuchstaben mit dem Ziel einer automatischen Formularerkennung. Schreibschrift und gedruckte Kleinbuchstaben werden damit nicht erkannt.

Sobald auf maschinenschriftlichem Schriftgut (z.B. Akten) handschriftliche Vermerke aufgebracht sind, tragen diese zu einer deutlich höheren Fehlerrate bei der Texterkennung bei, so daß eine OCR-Erfassung nicht mehr wirtschaftlich ist, im Vergleich zu einer händigen Textumsetzung.

Problematisch ist die Arbeit mit linguistischen Erkennungstechniken (Wörterbücher) bei der OCR-Bearbeitung von Archivgut, da hier moderne Sprachregeln und Orthographie auf ältere Sprachformen angewandt werden und somit Verfälschungen auftreten. So werden neue Fehler produziert.

Beim Training von Frakturschriften ist zu beachten, daß die Omnifont-Technologie in der OCR-Software abgeschaltet wird. Nach erfolgtem Zeichen-Training ist der trainierte Buchstabenbestand innerhalb der Software zu kontrollieren und eventuell nachzubearbeiten. Dies ist v.a. wichtig bei ähnlich geformten Buchstaben wie *f* und langem-*s* und bei Ligaturen.

Nach dem bisherigen Stand der Technik kann eine automatische Schrifterkennung aus wirtschaftlichen Überlegungen nur für solche Vorlagen empfohlen werden, die ein gutes Kontrastverhältnis und keine zusammenhängenden, zerstörten oder aufgebrochenen Buchstaben aufweisen. In allen übrigen Fällen erweist sich momentan die händige Textumsetzung durch Schreibkräfte als wirtschaftlicher.

Es hat sich allerdings gezeigt, daß die Texterkennung von filmdigitalisierten Vorlagen prinzipiell nicht schlechter ist als bei direktdigitalisierten Vorlagen, wenn bei der Erstellung der Scanfilme und der Filmdigitalisierung bestimmte Qualitätskriterien eingehalten werden.

Insgesamt ist die Entwicklung von Verfahren und Software für eine automatische Schrifterkennung noch in vollem Gange, so daß hier in nächster Zeit bessere Ergebnisse zu erwarten sind. Besonders die Erkennung von Handschriften, Symbolen und Fraktur-Schriften kann durch die Entwicklung und Weiterentwicklung spezieller mathematischer und linguistischer Verfahren verbessert werden.

Wirtschaftlichkeitsüberlegungen für die Digitalisierung von Archiv- und Bibliotheksgut

Von GERALD MAIER und PETER EXNER

Neben der Erarbeitung von Kriterien für die Qualitätskontrolle digitaler Images und die Ermittlung von technischen Parametern für die Bildaufbereitung wurden im Rahmen des Projekts *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Bücher und Archivalien* auch Marktanalysen und Wirtschaftlichkeitsüberlegungen im Bereich der Digitalisierung und hier insbesondere für die *Filmdigitalisierung* durchgeführt.

Kostenfaktoren für die Filmdigitalisierung

Die bei der Filmdigitalisierung entstehenden Kosten ohne Berücksichtigung der Kosten für Erschließung und Weiterverarbeitung der digitalisierten Formen sind von folgenden Faktoren abhängig:

- von der aktuellen Marktlage,
- von der Größe des Auftragsvolumens,
- von der verwendeten Scantechnik (u.a. Filmscanner, Aufsichtscanner),
- von der gewünschten Farbtiefe,[1]
- von der verwendeten Scanauflösung,
- und von der Vorlagenqualität (u.a. Qualität von Mikrofilmen).

Die Entscheidung für eine generelle Erhöhung der Auflösung beim Digitalisieren oder für die Wahl einer höheren Farbtiefe als 1 Bit beeinflußt mittelbar die Wirtschaftlichkeit der Konversion. Es entstehen höhere Datenmengen, die auch die Kosten für Datenlieferung, Datenspeicherung und Datenverarbeitung erhöhen.

Zusätzlich sind normalerweise noch die sog. Gemeinkosten für einen Auftrag und die Kosten für die Speicherung der Daten inklusive Datenträger und Transport zu berücksichtigen.

---

[1] Übliche Farbtiefen bei der Digitalisierung vom Mikrofilm sind 1 Bit (2 Farben, schwarzweiß), 4 Bit (16 Graustufen), 8 Bit (256 Graustufen) und 24 Bit (16,7 Mio. Farben).

Zu den nicht zu unterschätzenden Digitalisierungsgemeinkosten gehören

- die Kosten für die erstmalige Einrichtung des Filmscanners nach den Kundenanforderungen,
- das Rotieren von Images oder (manuelles) Ausschneiden der Bilder aus den Gesamtframes,
- eine eventuelle Indizierung der Images,
- und eventuell die damit verbundenen Programmierkosten.

Image Enhancement

Hohe, zusätzliche Kosten müssen auch für eine eventuell gewünschte oder notwendige individuelle Bildbearbeitung z.B. zur Bildverbesserung (*Image Enhancement*) eingeplant werden.[2]

Migration

Als zwangsläufiger Aufwand der Digitalisierung müssen auch die Folgekosten für die planmäßige Migration der Daten in die Wirtschaftlichkeitsbetrachtung einbezogen werden. Dabei kann es sich als wirtschaftlicher erweisen, im Bedarfsfall erneut vom vorhandenen Mikrofilm zu digitalisieren anstatt den Datenbestand regelmäßig zu migrieren.[3]

Marktlage

Die Kosten für das Digitalisieren sind zusätzlich von der Marktlage und der technischen Entwicklung abhängig, so daß zur Ermittlung der aktuellen Kosten immer wieder neue Marktanalysen notwendig sind.

Marktanalyse mit Kostenermittlung für Filmdigitalisierung

Für die Filmdigitalisierung wurde vom Projektteam eine Kostenermittlung bei mehreren Dienstleistern durchgeführt.[4] Dabei hat sich gezeigt, daß die Preisspanne beachtlich ist und u.a. aus dem Einsatz unterschiedlicher Scantechniken (automatische und manuelle Aufnahme, Einsatz eines Aufsicht-

---

[2] Zum *Image Enhancement* siehe in diesem Band den Beitrag von Gerald *Maier:* Qualität, Bearbeitung und Präsentation digitaler Bilder. S. 129.

[3] Zur Migrations-Thematik siehe in diesem Band den Beitrag von Hartmut *Weber*: Langzeitspeicherung und Langzeitverfügbarkeit digitaler Konversionsformen. S. 325.

[4] Vgl. dazu auch in diesem Band den Beitrag von Peter *Exner:* Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut. S. 113.

Übersicht über die Angebote für die Filmdigitalisierung (Preise in DM)

Angebot (Stamd)	*AHU*	*AHU*	*AHU*	*AHU*	*H*	*H*	*H*	*H*	*H*	*H*	*C*	*Makro*	*Makro*	*Makro*	*Makro*
Farbtiefe	*1 Bit*	*1 Bit*	*4 Bit*	*4 Bit*	*1 Bit*	*1 Bit*	*4 Bit*	*4 Bit*	*8 Bit*	*8 Bit*	*24 Bit*	*1 Bit*	*4 Bit*	*8 Bit*	*24 Bit*
Auflösung	*400 dpi*	*600 dpi*	*400 dpi*	*600 dpi*	*400 dpi*	*600 dpi*	*300 dpi*	*400 dpi*	*250 dpi*	*350 dpi*	*400 dpi*	*600 dpi*	*400 dpi*	*350 dpi*	*400 dpi*
1. (12/97)	0,29–0,63	0,29–0,63	0,29–0,63	0,29–0,63	0,75–2,88	0,75–2,88	0,75–2,88	0,75–2,88	0,75–2,88	0,75–2,88					
2. (12/97)	0,15 –0,40	0,30 –0,50	0,20–0,60	0,20–0,60	0,15 –0,40	0,30–0,50	0,20–0,60	0,20–0,60	0,25–0,85	0,25–0,85	2,50–25,00	15,00– 45,00	15,00– 45,00	15,00– 65,00	15,00–85,00
3. (12/97)	0,18	0,25	0,25	0,35	0,28	0,35	0,10	0,25	0,20	0,25	0,50	1,50	1,85	2,05	2,55
4. (12/97)	1,94	2,59	1,46	1,94	1,94	2,59	1,46	1,94	0,99	1,94	3,01	0,99	0,99	0,47	1,51
	3,45	4,49	3,04	3,45	3,45	4,49	3,04	3,45	2,42	3,45	6,39	2,42	2,42	1,73	3,71
5. (12/97)	0,35	0,52	0,81	1,15	0,35	0,52			0,81	1,15	4,60				
6. (07/98)	0,18–0,27	0,20–0,29	0,68–0,95		0,18–0,27	0,20–0,29	0,62–0,85	0,68–0,95	0,59–0,80	0,65–0,90	3,50–28,00	5,50–78,00			

Abkürzungen: AHU = Anti-Halation-Undercoated-Mikrofilm (panchromatisch); C = Farbmikrofilm; H =.Halbtonmikrofilm; Makro = Makrofiche.

scanners oder eines speziellen Filmscanners) und der damit oft verbundenen Qualitätsunterschiede resultiert. Außerdem bietet nicht jeder Dienstleister eine Digitalisierung in allen Auflösungen und Farbtiefen an. So können manche Dienstleister einen AHU-Mikrofilm mit 16 Graustufen nur über den Umweg von 256 Graustufen digitalisieren; dieser Scan muß dann auf 16 Graustufen heruntergerechnet werden. Umgekehrt können manche Dienstleister keine bitonalen Scans erzeugen, vielmehr müssen hier Graustufenscans in bitonale konvertiert werden.

Anmerkungen zur Übersicht über die Angebote für die Filmdigitalisierung

zu 1.:
- Vorgesehene Scannertechnologien: SunRise-Mikrofilmscanner
- Keine Angaben zu den Preisen für Digitalisierung von Farb-Mikrofilmen und Makrofiches
- Kosten für CD-R-Datenträger: DM 90,00 pro Stück

zu 2.:
- Vorgesehene Scannertechnologien: Film-/Dia-Scanner für 1 Bit, 8 Bit und 24 Bit; Durchlicht-Scanner; Trommelscanner
- Alle Preise verstehen sich als überschlägige Kostenangaben; lediglich die Preise für das Scannen von AHU-Filmen mit 400 dpi sind verbindlich.
- Die Preise für 35mm Farb-Mikrofilm unperforiert positiv beziehen sich auf 16, 64 und 256 Farben.
- Kosten für CD-R-Datenträger: DM 45,00 pro Stück
- Zusatzleistungen:
  - Daten-/Texterfassung nach Aufwand: DM 39,00/Stunde
  - Bildbearbeitung: DM 69,50/Stunde
  - automatische Texterkennung: DM 69,50/Stunde
  - Konvertierungen in die Formate HTML, SGML, PNG, JPEG, FlashPix: DM 69,50/Stunde

zu 3.:
- Vorgesehene Scannertechnologien: Mikrofilmscanner Bell und Howell D3000 und Fujix MS 6000
- Die Digitalisierung vom Farb-Mikrofilm erfolgt mit einem Durchlichtscanner
- Kosten für CD-R-Datenträger: DM 45,00 (Master), DM 35,00 (Kopie)

zu 4.:
- Vorgesehene Scannertechnologien: CCD-Aufsichtscanner mit einem Chip (4500 x 3500 Pixel), der bei der Aufnahme mit Piezo-Elementen bewegt wird.

- Die erste Zahlenreihe bezieht sich auf automatische Aufnahmen, die zweite auf manuelle. Beim automatischen Scannen werden im Gegensatz zum manuellen die Bildposition auf dem Mikrorollfilm oder dem Makrofiche computerunterstützt angesteuert. Bei Vorlagen, bei denen dies nicht machbar ist, werden die digitalen Bilder manuell erfaßt. Dies bedingt den höheren Preis.
- Bitonale Scans können nicht erzeugt werden, vielmehr werden die Graustufen-Scans in bitonale umgerechnet (deshalb gleiche Preise für bitonale und Graustufenbilder).
- Nicht in den Preisen enthalten sind sog. Regiearbeiten, die bei einem Projekt mit Kosten in Höhe von 15% der Scankosten zu Buche schlagen.
- Kosten für CD-R-Datenträger: DM 45,00 pro Stück

zu 5.:
- Vorgesehene Scannertechnologien: SunRise-Mikrofilmscanner
- Das Angebot für die Digitalisierung des AHU-Mikrofilms mit 16 Graustufen bezieht sich auf 256 Graustufen, da der verwendete Scanner generell Bilder mit dieser Farbtiefe erzeugt. Für die Digitalisierung des AHU-Mikrofilms mit 16 Graustufen müßten die generierten 256 Graustufen in einem zusätzlichen Arbeitsgang auf die geforderten 16 Graustufen heruntergerechnet werden.
- Der Preis bei der Digitalisierung des AHU-Mikrofilms mit 16 Graustufen von DM 0,70 bezieht sich auf ein Originalformat bis Doppelfolio, der Preis von DM 1,00 bezieht sich auf größere Formate.

zu 6.:
- Vorgesehene Scannertechnologien: keine Angaben
- Zusatzleistungen:
  - Indexieren von Attributwerten: DM 0,01–0,025 je Einzelzeichen
  - Automatische Texterkennung: DM 0,65–9,50/1000 Zeichen (ohne Korrektur und interaktives Training)
- Kosten für CD-R-Datenträger: DM 35,00 (Master); DM 28,00 (Duplikat)

Die Preise differierten zum Teil beachtlich. Die Digitalisierung eines AHU-Films, bitonal mit 600 dpi, kostete zwischen DM 0,20 und DM 4,49. Die Preise für digitale Bilder vom Halbtonfilm in 16 Graustufen mit 400 dpi lagen zwischen DM 0,20 und DM 3,45, ebenso bei 256 Graustufen und 350 dpi. Für die Digitalisierung von Farbfilmen, die nur noch fünf Dienstleister anbieten konnten, beliefen sich die Angebote bei gewünschten 400 dpi zwischen DM 0,50 und DM 28,00. Digitale Bilder von Makrofiches, die nur noch drei, bei farbigen Vorlagen vier Firmen realisieren konnten, kosteten bitonal mit 600 dpi zwischen DM 0,99 und DM 45,00, farbig und mit 400 dpi zwischen DM 1,51 und DM 85,00. Dabei waren die billigsten und teuersten Leistungen nicht durchgängig bei denselben Anbietern zu finden, sondern differierten.

Insgesamt schälten sich bei der Ausschreibung zwei Gruppen von Dienstleistern heraus:

1. Ein eingeschränktes Angebot von Dienstleistungen boten zwei Firmen: Die erste konnte digitale Dateien weder von Makrofiches noch vom Farb-Mikrofilm erzeugen. Die zweite konnte keine Makrofiches generieren.
2. Eine uneingeschränkte Leistungspalette boten drei Dienstleister, ein weiterer mit leichten Einschränkungen.

Zusammengefaßt reicht die Preisspanne (Stand: August 1998) je nach Auflösung und gewünschter Farbtiefe beim Scannen

- eines 35mm Films in 1 oder 8 Bit von ca. DM 0,20–3,50 pro Image,
- eines 35mm Films AHU-Films in 1 Bit von ca. DM 0,20–4,50 pro Image,
- eines 35mm Halbtonfilms in 8 Bit von ca.0,20–3,45 pro Image,
- eines 35mm Farbfilms in 24 Bit von ca. DM 0,50–28,00 pro Image und
- eines Farbmakrofiches von ca. DM 1,50–85,00 pro Image.

Dazu kommen normalerweise immer noch die Kosten für die Erstellung eines CD-R Datenträgers (incl. Rohling):

- Master: DM 35,00–90,00,
- Duplikat (falls angeboten): DM 28,00–35,00,

und die Kosten für Zusatzleistungen:

- Bildbearbeitung: DM 70–150/Stunde,
- Automatische Texterkennung: DM 70/Stunde,
- Konvertierungen: DM 70/Stunde,
- Vollständigkeitsprüfung: DM 60/Stunde.

Die Kosten für die Direktdigitalisierung bei einem Dienstleister

Bei der im Dezember 1997 durchgeführten Direktdigitalisierung reicht die Kostenspanne je nach Auflösung und gewünschter Farbtiefe

- bei einer manuellen Aufnahme von ca. DM 3,50–6,50 pro Image und
- bei einer automatischen Aufnahme von ca. DM 2,00–3,00 pro Image, wobei die Automatisierung aufgrund der Scantechnik hier nicht in dem Maße möglich ist wie bei der Filmdigitalisierung.

Kostenvergleich *Filmdigitalisierung – Direktdigitalisierung*

Vergleicht man die Kosten pro Image zwischen *Filmdigitalisierung und Direktdigitalisierung*, dann ergibt sich folgendes: Für ein Image mit 400 dpi Scanauflösung und 1 Bit-Farbtiefe betrugen sie im Dezember 1997 ohne Berücksichtigung der Gemeinkosten und individueller Bildverbesserung

- bei der Direktdigitalisierung ca. DM 2,00 (automatische Aufnahme)
- bei der Filmdigitalisierung ca. DM 0,50 (DM 0,25 AHU-Mikrofilmaufnahme, DM 0,25 Filmscan automatisch).

Dabei ergeben sich für die Direktdigitalisierung in 1 Bit-Farbtiefe und mit einer Auflösung von 400 dpi viermal höhere Kosten pro Image als bei der Filmdigitalisierung.

Bei den ermittelten Preisen ist zu bedenken, daß hier rasch Veränderungen auftreten können, so daß bei jeder größeren Auftragsvergabe eine erneute Preisermittlung notwendig ist.

Aus Gründen der Wirtschaftlichkeit sollte die Digitalisierung von Mikroformen grundsätzlich an externe Dienstleister in Form eines Outsourcings vergeben werden, da die hierzu notwendige technische Ausstattung und das Fachpersonal im eigenen Archiv kaum zur Verfügung stehen. Es ist auch nicht sinnvoll, große Investitionen im technischen Bereich zu tätigen, da hier mit sehr kurzen Innovationszyklen zu rechnen ist und somit sehr rasch die ganze Technik veraltet.

Bereitstellung und Nutzung von digitalisiertem
Archiv- und Bibliotheksgut

Kontextbezogene Online-Präsentationen von Archivgut

Von Karsten Uhde

Das Forschungsprojekt der Landesarchivdirektion Baden-Württemberg *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Bücher und Archivalien*, beschäftigte sich nicht nur mit der Digitalisierung von Archivalien und Bibliotheksgut selbst, sondern umfaßt darüber hinaus auch Untersuchungen zu Fragen der Präsentation der zuvor digitalisierten Quellen im Internet.

Die Arbeit einer kleinen Arbeitsgruppe der Archivschule, an der neben dem Autor noch drei Kursteilnehmer teilnahmen, zerfiel in drei größere Abschnitte:

1. In der ersten Phase der Arbeit wurde eine Bestandsaufnahme bereits vorliegender Archivalienpräsentationen erstellt und eine Analyse der Intentionen auf Anbieter- wie auf Nutzer-Seite durchgeführt.
2. In der zweiten Phase stand die Entwicklung eines tragfähigen Konzeptes zur Präsentation digitalisierter Quellen im Internet im Mittelpunkt.
3. Die Umsetzung dieses Konzeptes in verschiedene Präsentationen war Gegenstand des dritten Abschnitts, wobei in diesem Rahmen nur auf standardisierte, kontextbezogene Online-Präsentationen von Archivgut eingegangen wird.[1]

Bestandsaufnahme und Analyse der Intention auf Anbieter- und Nutzer-Seite

Zunächst wurde eine Bestandsaufnahme bereits vorliegender Archivalienpräsentationen erstellt und eine Analyse der Intentionen auf Anbieter- wie auf Nutzer-Seite durchgeführt.

---

[1] Zur Umsetzung des Fachkonzepts in verschiedene Formen von Präsentationen siehe auch in diesem Band die Beiträge von Thomas *Fricke*: Didaktische Internet-Präsentationen von Archivgut. S. 247, und den Beitrag von August *Wegmann*: Elektronische Bücher im Internet. S. 287.

Dabei wurden Internet-Angebote von Archiven, Bibliotheken, Museen etc. durchsucht, mit dem Ziel, Beispiele für die Präsentation digitalisierter Archivalien zu finden. Trotz einer intensiven Suche waren nur wenige wirklich aussagekräftige Beispiele zu finden, die noch dazu von sehr unterschiedlicher Qualität waren, so daß eine Analyse wenig Ergebnisse erwarten ließ.[2] Deshalb wurde die Suche nach Beispielen auf analoge Medien ausgeweitet, wobei hauptsächlich Publikationen von Archiven, Bibliotheken und Museen sowie Unterrichtsmaterialien herangezogen wurden.

Die anschließende Analyse erfolgte unter drei Fragestellungen:

1. Was will der Anbieter mit einer Präsentation erreichen?
2. Welche Intention hat der Nutzer?
3. In welcher Weise müssen die Quellen für den jeweiligen Zweck aufbereitet werden?

Aufgrund der Untersuchung der wenigen bereits bestehenden Präsentationen digitalisierter Archivalien im Internet, vor allem aber durch die der oben angesprochenen Kataloge, Broschüren, Arbeitsmappen und Faltblätter, in denen Quellen der Öffentlichkeit präsentiert werden, kam die Arbeitsgruppe zu dem Ergebnis, daß die Anbieter mit all diesen Produkten drei unterschiedliche Intentionen verfolgten.

1. Der Anbieter will die in seinem Besitz befindlichen Quellen dem Publikum präsentieren. Dies kann geschehen, um auf die Bedeutung der einzelnen Quelle aufmerksam zu machen (z.B. die Darstellung einer Stadtgründungsurkunde) oder auf die des gesamten Archivs und seiner Bestände.[3]
2. Der Anbieter will einem oder mehreren Nutzern Quellen zur Verfügung stellen, damit diese die Quellen dann selbständig auswerten können. Dabei kann dem Anbieter das genaue Thema der Nutzer schon vorher bekannt sein oder auch erst im Nachhinein bekannt werden.[4]

---

[2] Inzwischen hat sich in diesem Bereich einiges verändert. Viele Archive und vor allem amerikanische Bibliotheken zeigen inzwischen digitalisierte Archivalien. Das Fazit, daß diese Präsentationen von sehr unterschiedlicher Qualität sind, ist allerdings noch immer zutreffend.

[3] Rein illustrativen Charakter hat beispielsweise die Darstellung des Grundrisses der Stadt Nürnberg von Süden um 1630 auf einer Seite des Stadtarchivs Nürnberg (URL: *http://www.stadtarchiv.nuernberg.de/1.htm*).

[4] Ein Beispiel für diese Form der teilweise massenhaften Darstellung sind die vom Stadtarchiv Duderstadt im Rahmen eines Forschungsprojektes in den Jahren 1996–1998 digitalisierten und im Internet abrufbaren mittelalterlichen Amtsbücher (URL: *http://www.archive.geschichte.mpg.de/duderstadt/ab/rep10.htm*). Auch die Plakatsammlung des Archivs für soziale Demokratie in Bonn würde noch in diesen Bereich fallen (URL: *http://www.fes.de/archive/galerie/plak_index.html*).

3. Der Anbieter benutzt die Quellen, um ein bestimmtes, zumeist historisches Thema der Öffentlichkeit vorzustellen. Im Gegensatz zu den beiden erstgenannten Zielen, die vor allem bei Archivaren anzutreffen sind, sind die meisten Anbieter, die diese Intention verfolgen, Historiker, Lehrer, Mitarbeiter von Museen oder Archivpädagogen, die oft – aber nicht immer – mit Zustimmung und im Auftrag des Archivs oder der Bibliothek handeln.

Der zentrale Unterschied dieser drei Intentionen liegt im Umfang der Aufbereitung. Während in den beiden zuerst genannten Fällen keine oder nur verhältnismäßig wenig Erklärungen zu den einzelnen Quellen geliefert werden, sind bei der dritten Form in der Regel umfangreiche Aufbereitungen, wie z.B. Transkriptionen und Erläuterungen zum historischen Hintergrund der Quelle, die Regel.[5]

Die zweite Fragestellung galt den Absichten und Bedürfnissen der Nutzer. Diese sind im Gegensatz zu den Zielen der Anbieter oft schwer zu fassen und nur indirekt dadurch zu ermitteln, daß ihre Absichten und Bedürfnisse bei einer erfolgreichen Präsentation mit den Absichten der Anbieter korrespondieren müssen. Auch hier kann man drei Typen von Nutzern unterscheiden:

1. Der Nutzer, der sich lediglich an der Präsentation der Quellen erfreut und dabei nur die formale bzw. ästhetische Komponente der Präsentation im Sinn hat. Dabei ist die Haltung des Nutzers rein rezeptiv. Eine intensive, selbständige Auseinandersetzung mit den Quellen wird nicht angestrebt, ebensowenig wie eine inhaltliche Auswertung derselben.
2. Der Nutzer, der Quellen für eigenständige Untersuchungen eines von ihm frei gewählten Themas nutzen will. Dabei handelt es sich um den klassischen Archivbenutzer. Dieser ist aktiv an der Aufbereitung und Auswertung der Quellen beteiligt, benötigt dazu allerdings einige grundlegende Informationen des Anbieters, wie z.B. die Signatur des Stücks.
3. Der Nutzer, der sich über ein Thema informieren will, daß ihm von einer anderen Person oder Institution vorgestellt wird. Ein typischer Vertreter dieser Gruppe ist der Besucher historischer Ausstellungen. Dieser Nutzertyp rezipiert die Auswertung und Aufbereitung der Quellen durch Dritte.

---

[5] Eines der ältesten, leider bisher nicht zu Ende geführten Beispiele für diese Art der Präsentation im Internet, ist die Präsentation der Urkunden des Passauer Stadtarchivs. Neben dem Faksimile der jeweiligen Urkunde sind hier Transkriptionen, zum Teil aber auch Erklärungen zu einzelnen Personen und zum historischen Hintergrund der Urkunde abrufbar (URL: *http://www.stadtarchiv-passau.de*).

Daraus resultiert, daß der Umfang der Aufbereitung der Quellen auch aus Sicht der Nutzer sehr unterschiedlich sein sollte. Während sich – trotz ihrer unterschiedlichen Intention – die Benutzer der beiden ersten Typen mit geringen Zusatzinformationen zufrieden geben, umfangreichere Darstellungen unter Umständen sogar als störend empfinden können, bevorzugt der Nutzer im dritten Fall eine möglichst ausführliche Aufbereitung.

Zusammengefaßt ergeben sich auf die Frage, wie umfangreich die Zusatzinformationen bei der Präsentation der digitalisierten Archivalien sein sollten, zwei Antworten:

1. Für die hauptsächlichen Bedürfnisse sowohl von Anbieter- als auch von Nutzer-Seite ist eine verhältnismäßig geringe Menge von Zusatzinformationen vollkommen ausreichend.
2. Für bestimmte, eher im Bereich der politischen Bildungsarbeit und des Ausstellungswesens liegende Anliegen sind umfangreiche Zusatzinformationen und Aufbereitungen der Quellen notwendig.

Fachkonzept zur formalen und inhaltlichen Aufbereitung der unterschiedlichen Gattungen von Archiv- und Bibliotheksgut für den Zugriff im Internet

In Kooperation mit der Archivschule Marburg wurde ein Fachkonzept zur formalen und inhaltlichen Aufbereitung der unterschiedlichen Gattungen von Archiv- und Bibliotheksgut für den Zugriff im Internet erarbeitet.

Das Fachkonzept beinhaltet

- Beispiele der formalen und inhaltlichen Aufbereitung der unterschiedlichen Gattungen von Archiv- und Bibliotheksgut. Hierbei wurden sowohl konventionelle Formen der Präsentation von Archivalien und Büchern als auch Präsentationen im Internet erfaßt. Aus diesen Beispielen wurden positive wie negative Merkmale herausgefiltert und die dabei gewonnenen Erkenntnisse zusammengestellt.
- Analyse des Bedarfs von tatsächlichen und potentiellen Nutzern, die Angebote digitaler Konversionsformen von Archivalien und Büchern nutzen oder nutzen wollen. Ausgehend von den konventionellen Präsentationsformen wurden die erweiterten Bedürfnisse tatsächlicher und potentieller Benutzer ermittelt, die sich durch eine Internet-Präsentation ergeben. Zudem wurden Wege ermittelt, wie Interessenten, die bislang keine Archivbesucher waren, durch das Medium Internet als Benutzer angesprochen werden können.

- Festlegung eines Pflichtenheftes für die Programmierung solcher Modelle.

Typen von Präsentationsformen für Archiv- und Bibliotheksgut

Nach der durchgeführten Bestandsaufnahme und Analyse kann zwischen drei verschiedenen Typen von Präsentationsformen für Archiv- und Bibliotheksgut unterschieden werden:

Typ A: Die Zimelienschau

Der Archivar oder Bibliothekar will hierbei die Schätze seines Hauses einer breiten Öffentlichkeit darbieten. Er wird dabei in der Regel keine eingehendere Beschäftigung des Nutzers mit diesen konkreten Quellen beabsichtigen (wegen ihrer Besonderheit und damit ihres Wertes oder aus Gründen der Bestandserhaltung manchmal wohl sogar ausdrücklich ausschließen wollen), sondern maximal eine allgemeine Steigerung des Wissens der Nutzer über die Quellen und damit auch über die diese Quellen aufbewahrende Institution. Diese Präsentationsform spricht diejenigen Nutzer an, die sich an der formalen oder ästhetischen Besonderheit von einzelnen Quellen erfreuen wollen. Die Haltung des Nutzers ist rein rezeptiv; die Präsentation wird zum Kunstgenuß. Die Archivalien und Bücher werden in der Regel nur auszugsweise präsentiert und inhaltlich nur unzureichend bearbeitet.

Typ B: Die Präsentation verschiedener Quellen zur Darstellung eines Themas

Der Archivar oder Bibliothekar will (oder muß!) ein konkretes, zumeist historisches Thema darstellen und nutzt eine Reihe der ihm zur Verfügung stehenden Quellen, um dieses Ziel zu erreichen. Das Einzelstück wird hierbei in einen thematischen Interpretationsrahmen gestellt, der vom Anbieter festgelegt wird, und dem der Nutzer folgen muß. Dies entspricht der Intention des Nutzertyps, der die Präsentation konkret als Erweiterung seines Wissens über ein bestimmtes Thema nutzt, was eine gleichzeitige Betrachtung eines einzelnen Stücks als Zimelie allerdings nicht verhindert. Seine Beschäftigung mit den Quellen ist wie schon im ersten Fall weitgehend rezeptiv, kann allerdings im Einzelfall dazu führen, daß der Nutzer so viel Interesse entwickelt, daß er sich selbst aktiv und zumeist in vertiefender Form mit einzelnen der präsentierten Quellen auseinandersetzen will und damit nach der dritten Präsentationsform fragt. In Hinblick auf die Quelle bedeutet dies, daß sie häufig nur ausschnittsweise gezeigt wird und ihre Aufbereitung zumeist nur in eine bestimmte Richtung erfolgt.

Typ C: Die Präsentation einer Quelle zur Nutzung für verschiedene Themen

Bei der dritten Form präsentiert der Archivar oder Bibliothekar dem Nutzer eine einzelne Quelle ohne eine konkrete Auswertung. Dieser hat dann die Möglichkeit, mit frei zu wählender Methode, die Quelle in verschiedener Hinsicht auszuwerten und zu interpretieren. Die Auswahl der Quellen kann durch den Anbieter erfolgen, der sie potentiellen Nutzern vorstellen möchte, weil er sie grundsätzlich für interessant erachtet. Die Präsentation kann aber auch im Auftrag des oder der späteren Nutzer erfolgen. Der Anbieter wird die Quelle jedoch in einigen Fällen aufbereiten müssen, je nach Schwierigkeitsgrad der Quelle und nach Kenntnisstand und Absichten des Nutzers. Sind der Nutzer und sein Thema bei der Auswahl der Quelle nicht bekannt, so muß der Anbieter – selbstverständlich immer unter Beachtung des damit verbundenen Zeit- und Arbeitsaufwandes – den Grad der Aufbereitung recht hoch ansetzen, um möglichst vielen der möglichen Nutzungsvarianten gerecht zu werden.

Bindeglied zwischen Anbieter und Nutzer sind in allen Fällen die Quellen, die ebenfalls in formaler wie inhaltlicher Hinsicht in zahlreiche Gruppen zerfallen. So weist allein der *große Warenkorb* 16 verschiedene Quellengattungen auf, die teilweise noch weiter untergliedert sind.

Als Ergebnis kann festgehalten werden, daß die Auswahl einer Präsentationsform den Benutzerkreis festlegt. Deshalb wurden im Anschluß an das Fachkonzept verschiedene Formen von Präsentationsmodellen entwickelt, um neben den verschiedenen Gattungen von Archiv- und Bibliotheksgut auch verschiedenen Benutzerkreisen gerecht zu werden.

Pflichtenheft für die formale und inhaltliche Aufbereitung von Präsentationen für Archiv- und Bibliotheksgut

Bei der Entwicklung von Präsentationsmodellen für Archiv- und Bibliotheksgut sollte grundsätzlich folgendes berücksichtigt werden:

- lesbare Größe der Objekte,
- Identifikation der Objekte (Fundort, Signatur),
- formale Angaben (Metadaten)
- normale Navigation (Vor- und Zurückblättern) und
- *Mehrwert* gegenüber einer konventionellen Präsentation.

Der *Mehrwert* gegenüber einer konventionellen Präsentation und der Nutzung von Archivalien und Bibliotheksgut kann bei Online-Präsentationen unter anderem durch folgendes erzielt werden:

- erweiterte Zugänglichkeit unabhängig von Ort und Zeit,
- einen mehrdimensionalen Zugriff,
- erweiterte Navigationsmöglichkeiten,
- Erschließung auf Dokumentenebene im Kontext,
- Integration in Tektonik und Kontext (Kommentare, didaktische Aufbereitung) und
- Lesehilfen.

Im einzelnen wurden folgende technische Rahmenbedingungen für Präsentationsmodelle festgelegt:

- Verwendung von Internet-Technologie (HTML, DHTML, Java Script). Grundlage der Programmierung sollte HTML in der Version 4.0 sein.
- keine Verwendung von Java, Aktive-X oder anderer proprietärer Programmiersprachen
- Browser-Versionen: Netscape-Navigator/Communicator 4.x und Microsoft Internet Explorer 4.x.
- Bildschirmauflösung: 800x600 Pixel, in Zukunft 1024x768 Pixel
- Dateiformate für Bildschirmdarstellung: JPEG, in Zukunft auch PNG- und Wavelet-Formate

Standardisierte Präsentationsmodelle für Archivgut

Neben den individuell gestalteten Präsentationsmodellen wurden in Kooperation mit der Archivschule Marburg als Umsetzung des erstellten Fachkonzepts (Typ C: Die Präsentation einer Quelle zur Nutzung für verschiedene Themen) standardisierte Internet-Präsentationsmodelle erstellt. Verschiedene Beispiele für die Präsentation verschiedener Gattungen von Archivgut wie Amtsbücher, Akten, Urkunden, Siegel, Karten, Wappen und Photographien sind hier unter einer gemeinsamen Auswahlseite vereinigt worden. Die gesamte Präsentation ist im Internet jederzeit abrufbar.[6]

Umsetzung des Fachkonzepts

Um verschiedene Ansichten auf die Objekte zu gewährleisten, wurden in der Regel von jedem Objekt durch Aufbereitung der digitalen Master (Format-

---

[6] Die Präsentationen der Marburger Arbeitsgruppe finden sich unter: *http://www.lad-bw.de/digpro/texte/uhde/ladhome.html.*

konvertierung, Verkleinerung, Schärfen, Tonwertkorrektur) drei verschiedene Bild-Fassungen erzeugt:

1. Ein Vorschaubild (Thumbnail) der Quelle bzw. einer einzelnen Seite. Diese soll dem Nutzer vor allem einen allerersten Eindruck bei mehrteiligen Quellen vermitteln. Die Dateien haben zumeist eine Dateigröße zwischen 5 und 20 KB und benötigen nur eine kurze Ladezeit.
2. Eine bildschirmfüllende Ansicht des Objekts (Vollbild), um dem Betrachter einen Gesamteindruck zu ermöglichen. Diese Dateien sind meist zwischen 30 und 120 KB groß.
3. Eine Detailansicht, die es dem Betrachter erlaubt, die Schrift und bildliche Darstellungen einwandfrei zu erkennen. In einigen Fällen könnten die beiden letztgenannten Fassungen identisch sein. Diese Dateien sind zwischen 100 und 250 KB groß, wobei in einzelnen Fällen die Dateigröße noch darüber hinausgehen mußte, um eine Lesbarkeit zu gewährleisten.

Über die technischen Rahmenbedingungen hinaus wurde festgelegt, welche Zusatzinformationen bei allen Präsentationen unbedingt anzugeben sind:

1. Fundort und Signatur der Archivalien, u.U. auch Angaben zur Lage innerhalb einer Einheit.
2. Formalbeschreibung: sie enthält alle Angaben, die aufgrund der Darstellung im Internet nicht mehr ermittelbar sind, wie z.B. die Größe des Archivale, Angaben zu seinem Material, zu eventuellen Restaurierungen/Beschädigungen usw.
3. Technische Angaben zu den Bilddateien: dazu zählen Angaben zum Datei-Format, zur Dateigröße, aber auch zu den verwendeten Programmen und Bearbeitungsschritten, mit denen die Ursprungsdatei in die präsentierte überführt wurde.

Um sicher zu stellen, daß diese zumindest rudimentären Angaben zu den einzelnen Archivalien schon zu Beginn der Arbeiten erhoben werden, wurde ein Formular entwickelt, in das die wichtigsten Merkmale aufgenommen werden können. Dieses Formular begleitetet die Archivalien durch alle Bearbeitungsschritte.

Nummer: **Mikrofilm Nr.:**

Archiv/Bibliothek: .....

Signatur: .....

Titel: .....

.....

Umfang der Archivalie insgesamt:

..... Seiten / Blätter / Bögen / Bände

Aufgenommene Seiten/Blätter/Bögen: .....

.....

.....

Material:

Pergament: O

Papier: O

Wachs: O

Sonst.: .....

.....

Einband: .....

.....

.....

Datei	Dateiname	Datei-format	Größe i. KB	% von Urdatei	Bearbeitungsschritte
Detailbild					
Vollbild					
Übersichtsbild					
...					
...					
...					

Sonstige Bemerkungen:

.....

.....

.....

.....

.....

.....

.....

Abb. 1: Formular für die Archivaliendigitalisierung.

Das grundlegende Problem jeder Präsentation ist jedoch nicht die Erhebung dieser Daten, sondern die Verbindung der digitalisierten Bildobjekte mit den entsprechenden Erschließungsdaten und sonstigen Metadaten.

Abb. 2: Beispiel für die Dreiteilung des Gesamt-Bildschirms.

Für die Basispräsentationen wurde der Bildschirm in drei Fenster (Frames) unterteilt:

- Das Mittlere, daß ca. 70% der gesamten Höhe und 100% der Gesamtbreite einnimmt, zeigt die Quelle.
- Das obere beinhaltet Lagerungsort, Signatur und einen Kurztitel, dient also der Orientierung, um welche Quelle es sich handelt. Zudem wird hier mit Hilfe kleiner Buttons die Navigation zwischen den verschiedenen Größen eines Bildes und unter Umständen auch zwischen einzelnen Seiten/Teilen umfangreicherer Einheiten ermöglicht.
- Das untere Feld dient der Aufnahme von Buttons, mit denen die Formalbeschreibung und technische Angaben zu jedem Objekt eingeblendet werden können. Zusätzlich kann über diese Leiste zwischen den verschiedenen Teilen einer Quelle navigiert werden.

Während die Quelle auf dem Bildschirm betrachtet wird, können zusätzliche Fenster mit technischen Hinweisen und einer Formalbeschreibung geöffnet werden.

Abb. 3: Beispiel für eine Seite mit geöffnetem Zusatzfenster.

Für Archivalien, die aus mehreren Teilen bestehen, wurde zusätzlich zu den Navigationsleisten die Funktion des *Vor- und Zurückblätterns* über zwei Buttons realisiert.

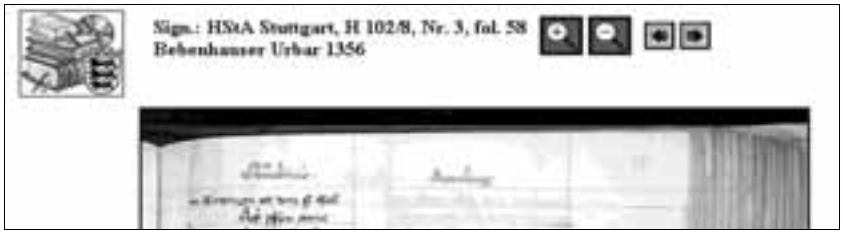


Abb. 4: Navigationsleiste mit Buttons zum Vor- und Zurückblättern.

Karten werden so gestaltet, daß Legenden oder erklärende Texte jederzeit in Extrafenstern nutzbar sind.

Bei Notariatsinstrumenten wird das Signet, bei Siegelurkunden die Siegel gesondert dargeboten. Wenn das Siegel Stempel auf Vorder- und Rückseite aufweist, werden beide Seiten dem Betrachter geboten.

Bei Photos und Plakaten werden neben dem Titel auch die Namen der Photographen, Designer etc. aufgeführt sowie der Tag der Veröffentlichung sofern die Angaben bekannt sind.

Abb. 5: Beispiel für die Verwendung eines zusätzlichen Navigationsframes.

Wenn ein Stück aus mehr als zehn Einzelteilen besteht, ist eine Navigation über den unteren Frame nicht mehr sinnvoll. Hier wurde ein zusätzliches Fenster in die Präsentation eingefügt, über das die Navigation erfolgen kann.

Bei Siegeln wird die Möglichkeit der Drehung um zumindest jeweils 45° angeboten, damit die Umschrift jederzeit ohne Schwierigkeiten zu lesen ist.

Evaluation und Zusammenfassung

Im Hinblick auf eine Qualitätskontrolle und Akzeptanz der Internetpräsentation wurde durch den 35. Fachhochschulkurs der Archivschule Marburg eine Evaluation durchgeführt.

Dabei wurden vor allem die verhältnismäßig langen Ladezeiten für die Vollbilder, hauptsächlich aber für die Detailansicht der Archivalien häufig hervorgehoben. Zugleich wurde aber bescheinigt, daß die einzelnen Seiten in dieser Ansicht sehr gut lesbar seien. Im Vergleich mit der Benutzung der Originale wurden in fast allen Fällen keinerlei Einbußen bei der Bildschirmarbeit festgestellt.

Obwohl die Basispräsentationen grundsätzlich keinerlei didaktische Aufbereitungen bieten sollten, wurde von den meisten Kolleginnen und Kollegen angemerkt, daß einzelne über die normale Präsentationsgestaltung hinausgehende Zusätze die Nutzung durchaus noch weiter verbessern könnten. Als ein Beispiel wurde die Möglichkeit von Drehungen beim Siegel genannt, die die Betrachtung und auch Beschreibung des Siegels überhaupt erst möglich macht.

Insgesamt läßt sich für die standardisierte Archivalienpräsentation folgendes festhalten:

1. Die Erstellung einer Präsentation ist – solange sie individuell erfolgen muß – sehr zeitaufwendig. So dauerte die Aufbereitung einer nur aus einem Bild bestehenden Einheit zunächst ca. 70 Minuten; wenn bereits bestehende Dateien nur umgeschrieben werden mußten, waren noch ca. 30 Minuten zu veranschlagen. Dabei wurde für die Bildaufbereitung der digitalen Master zwischen 10 und 15 Minuten pro Objekt benötigt. Durch Automatisierungen sind hier erhebliche Zeiteinsparpotentiale möglich.
2. Es hat sich herausgestellt, daß selbst bei umfangreicheren, mehrteiligen Archivalien fast jede Seite einzeln bearbeitet werden muß. Das bedeutet, daß die einzelnen Schritte der Bearbeitung laufend dokumentiert werden sollten, zumal sie im Nachhinein kaum noch nachzuvollziehen sind. Hierbei hat sich die Verwendung des entwickelten Formulars als sinnvoll herausgestellt.
3. Die Präsentation besonders großformatiger Stücke (Notariatsinstrumente) ist im Internet auf herkömmliche Weise kaum möglich, da die Dateien für den Transfer im Netz zu groß sind. Hierfür müssen besondere technische Lösungen entwickelt und angewandt werden, wie sie schon für die Präsentation von Zeitungen vorliegt.[7] Dies erfordert allerdings einen erheblichen Mehraufwand bei der Erstellung der Präsentation. Ob dieser in allen Fällen sinnvoll ist, bleibt im Einzelfall zu klären.

Zusammenfassend ist festzustellen, daß sich das Konzept als tragfähig erwiesen hat. Allerdings müssen für eine rationelle Erstellung von Standardpräsentationen die Modelle noch auf eine Massenverarbeitung hin optimiert werden, was im Rahmen dieses Projektes nicht zu leisten war.

---

[7] Siehe dazu in diesem Band den Beitrag von Thomas *Fricke*: Internet-Präsentationsmodell für Zeitungen und Großformate. S. 297.

Didaktische Internet-Präsentationen von Archivgut

Von Thomas Fricke

Ein wichtiges Ziel des Projekts *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Bücher und Archivalien* war die Präsentation von digitalisiertem Archiv- und Bibliotheksgut im Internet. Unter Berücksichtigung von Benutzerbedürfnissen und archivfachlichen Gesichtspunkten wurden verschiedene, didaktisch aufbereitete Präsentationsmodelle anhand ausgewählter Archivalien entwickelt, die die mit der Internet-Technologie verbundenen neuen Möglichkeiten eines mehrdimensionalen Zugriffs nutzen.

Kernstück einer jeden Präsentation sind die Quellenseiten. Die Begleittexte wurden bewußt kurz gehalten. Die Archivalien werden immer in mehreren Ansichten dargeboten: zunächst in einer der Bildschirmgröße angepaßten Gesamtansicht und dann in vergrößerten Ansichten, die das Lesen der Schriften und Erkennen der Details ermöglichen. Von der Gesamtansicht der Quelle kann der Benutzer häufig durch einen einfachen Klick mit dem Mauszeiger zu einer vergrößerten Ansicht oder einem vergrößerten Ausschnitt gelangen.

Die Präsentationsmodelle beinhalten eine Transkriptionshilfe für handschriftliche Archivalien. Beim Überfahren oder Berühren der handschriftlichen Originalzeilen mit dem Mauszeiger wird der Cursor zur Hand, und es wird in einem Anzeigefeld die Umschrift eingeblendet. Der Benutzer kann sich durch die Präsentation führen lassen und bei Bedarf eine Hilfefunktion aufrufen. Inhaltsübersicht, Index und Literaturverzeichnis unterstützen ihn bei der Arbeit mit der Quelle.

Alle Modelle können online im Internetangebot des Projektes betrachtet werden *(http://www.lad-bw.de/digpro/praes.htm*).

Modell 1: Prozeß gegen Ludwig Pfau 1848/49

Das erste Präsentationsmodell behandelt die Zensur in der Revolutionszeit 1848/49 im deutschen Südwesten, insbesondere den Zensur-Prozeß gegen den Redakteur des politisch-satirischen Wochenblatts *Der Eulenspiegel* Ludwig Pfau. Ein Beitrag von Ulrich Maier in der Beilage *Quellenmaterial für den Unterricht* der von der Landesarchivdirektion herausgegebenen Archivnachrichten Nr. 13 vom November 1996 bildet die inhaltliche Grundlage der Internetanwendung.

Abb. 1: Internet-Präsentationsmodell *Zensur in der Revolutionszeit 1848/49 (http://www.lad-bw.de/digpro/texte/fricke1/eulpau.htm).*

Von der mit einer einladenden Illustration versehenen Eröffnungs- oder Titelseite gelangt der Benutzer zur Einführungsseite, auf der die historische Thematik der Präsentation kurz umrissen wird. Der Bildschirm ist ab hier dreigeteilt in Hauptfenster, Auswahl- und Fußleiste. Das Hauptfenster zeigt die über die Auswahlleiste anwählbaren Inhalte an. Die Auswahlleiste selbst enthält vier Hauptauswahlpunkte: Inhaltsübersicht, Themen, Quellen und Hilfen. Zu den drei letzten sind wieder zahlreiche Untermenüpunkte vorhanden. In der Fußleiste befindet sich ein Anzeigefeld. Hier werden Erläuterungen zu dem im Hauptfenster dargestellten Inhalt eingeblendet, sobald der Mauspfeil auf ein entsprechendes Anmerkungssymbol im Hauptfenster trifft. Es gibt zwei solcher Anmerkungssymbole: einen schwarzen Pfeil, durch den erklärungsbedürftige Namen und Begriffe gekennzeichnet sind, und ein umrahmtes *Q*, das bei Zitaten oder Quellendarstellungen steht und hinter dem sich Quellennachweise verbergen.

Rechts neben dem Anzeigefeld befindet sich ein Fragezeichen-Symbol zum Öffnen eines frei verschiebbaren Extrafensters mit der Online-Hilfe. Sie beschreibt die einzelnen Funktionen und Navigationssymbole der Präsentation. Ist die Einführungsseite angewählt, weist ein Lauftext in der Statuszeile des Browsers, dem grauen Feld unten am Bildschirm, auf die Hilfefunktion hin.

Neben der Möglichkeit einzelne Inhalte der Präsentation auszuwählen, kann der Benutzer sich mithilfe eines graphischen Symbols auch durch die Präsentation führen lassen und so alle Themen- und Quellenseiten in der inhaltlich richtigen Reihenfolge passieren. Rechts oben im Hauptfenster befindet sich eine Statusanzeige, die den jeweiligen Stand der Führung zeigt. In der Statuszeile des Browsers wird das jeweilige nächste Ziel genannt.

Abb. 2: Internet-Präsentationsmodell *Zensur in der Revolutionszeit 1848/49.* Dreiteilung des Bildschirms, Karikatur *Romanus Triumphator* aus dem *Eulenspiegel* vom Juni 1849 mit Abtastfunktion.

Es werden Themen- und Quellenseiten unterschieden, die durch spezielle Hintergundfarben gekennzeichnet sind. Die Quellenseiten, die den Kern der Präsentation bilden, sind in drei Teile gegliedert: Einführungstext, Originalquelle mit Transkriptionshilfe und Umschrift mit Erläuterungen. Eine besondere Hilfe für den Benutzer beim Online-Lesen der handschriftlichen oder in Fraktur gesetzten Quellentexte stellt die bereits beschriebene Transkriptionshilfe dar. Beim Überfahren der einzelnen Zeilen mit dem Mauspfeil wird im Anzeigefeld die Umschrift der entsprechenden Textzeile angezeigt. Zusätzlich kann ein frei verschiebbares Extrafenster geöffnet werden, das eine Tafel mit dem Alphabet der Sütterlinschrift enthält. Außerdem steht eine zusammen-

hängende Umschrift der Quelle mit aktenkundlichen und inhaltlichen Erläuterungen für eine Interpretation der Quelle zur Verfügung. Durch verschiedenfarbige Balken und aktenkundliche Bezeichnungen am Rand des Textes wird die inhaltliche Gliederung des Schriftstücks verdeutlicht.

Die bei der Transkriptionshilfe eingesetzte Technik eignet sich auch hervorragend zur Erläuterung von Bildquellen. Beim Abtasten einer Grafik mit dem Mauszeiger, können entsprechende Angaben zum Inhalt des Bildes im Anzeigefeld eingeblendet werden.

Die in der Ausgabe der Zeitschrift *Eulenspiegel* vom Juni 1849 enthaltenen Artikel und Grafiken können sowohl über eine Titelliste als auch visuell durch Anklicken von Bereichen innerhalb eines Vorschaubildes angewählt werden (mehrdimensionaler Zugriff). Beim Überfahren der einzelnen Komponenten der Zeitschrift innerhalb des Vorschaubildes wird im Anzeigefeld der jeweilige Titel angezeigt.

Darüber hinaus ist ein umfangreicher Sach- und Personenindex vorhanden. Über anklickbare Schlagworte kann direkt die entsprechende Seite aufgerufen werden. Einem besseren Verständnis der Quelle und ihrer Einordnung in den Kontext der Zeit dienen ein Literaturverzeichnis und eine Auswahl von Links für die weitere Materialsuche zum Thema.

Modell 2: Königsurkunde für das Kloster Gottesaue 1110

Das zweite Präsentationsmodell widmet sich der Stiftungsurkunde für das Benediktinerkloster Gottesaue aus den Beständen des Generallandesarchiv Karlsruhe. Diese Urkunde König Heinrichs V. von 1110 ist der älteste Beleg für die Besiedlung des heutigen Karlsruher Stadtgebiets und eine wichtige Quelle für die Frühgeschichte des Karlsruher Raumes. Das Präsentationsmodell basiert inhaltlich auf einem Beitrag von Stefan Schipperges in der Beilage *Quellenmaterial für den Unterricht* der Archivnachrichten Nr. 12 vom Mai 1996.

Der Grundaufbau dieses Modells entspricht dem vorigen. Im Mittelpunkt der Präsentation steht auch hier die archivalische Quelle. Die Gesamtansicht lädt zum selbständigen Erforschen der Urkunde ein. Berührt der Mauszeiger eine neben dem rechten Rand der Quelle befindliche Linie, so werden bestimmte Bereiche auf der Urkunde durch Schattierungen markiert und umrahmt, so daß der Aufbau der Quelle erkennbar wird. Beim Überfahren der markierten Bereiche mit dem Mauszeiger werden im Anzeigefeld Angaben zu den einzelnen Komponenten der Urkunde und dem Erhaltungszustand der Archivalie eingeblendet.

Die Detailansicht bietet lesbare Abschnitte der Urkunde. Beim Überfahren der einzelnen Zeilen wird im Erläuterungsfeld die Umschrift der entsprechenden Passage angezeigt. Ein Klick auf eine Urkundenzeile bewirkt, daß der Urkundenabschnitt verschwindet und sofort eine Gegenüberstellung von

Umschrift und Übersetzung genau dieser angeklickten Zeile erscheint. Zusätzlich läßt sich ein separates Hilfefenster mit einem Alphabet der sog. Diplomatischen Minuskel und weiterer Schriften einblenden, das auch weiterführende Literaturangaben zur Schriftenkunde enthält.

Eine weitere Quellenseite zeigt den lateinischen Urkundentext in seiner Gesamtheit und diesem tabellarisch gegenübergestellt die deutsche Übersetzung. Erklärungsbedürftige Stellen im Text sind hier ebenfalls mit schwarzen Pfeilen gekennzeichnet, so daß beim Berühren des Symbols mit dem Mauszeiger im Anzeigefeld Erläuterungen angezeigt werden. Ein Bild von der inhaltlichen Struktur des Urkundentextes kann sich der Benutzer anhand der letzten Quellenseite machen. Die einzelnen Quellenabschnitte sind mit den entsprechenden lateinischen Fachbezeichnungen versehen, deren Bedeutung wiederum im Anzeigefeld erklärt wird.

Auf den Themenseiten der Präsentation findet der Benutzer auch eine Karte der Besitzungen des Klosters Gottesaue nach der Urkunde von 1110. Zeigt man mit dem Mauspfeil auf einen Ortsnamen dieser Karte, so wird dieser im Anzeigefeld mit einer näheren Bestimmung angezeigt. Ein umfangreicher Orts- und Personenindex, Literaturangaben sowie eine Auswahl von Links zum Thema vervollständigen das Modell.

Modell 3: Der Fall des jüdischen Rechtsanwalts Waldeck 1933

Das dritte Präsentationsmodell thematisiert die Entwürdigung und Entrechtung der Juden im nationalsozialistischen Deutschland. Anhand badischer Akten werden die Auseinandersetzungen um die Zulassung des Mannheimer Rechtsanwalts Dr. Florian Waldeck zur Rechtsanwaltschaft im Jahre 1933 dargestellt. Der Vizepräsident des badischen Landtags und DVP-Fraktionsvorsitzende Waldeck verlor bei der Machtübertragung auf die Nationalsozialisten seine politischen Ämter und sah sich in den Folgejahren zunehmender Unterdrückung und Bedrohung ausgesetzt. Grundlage des Präsentationsmodells ist ein Beitrag von Hans-Georg Merz in der Beilage *Quellenmaterial für den Unterricht* der Archivnachrichten Nr. 6 vom Mai 1993.

Vom Aufbau her entspricht das Modell den beiden vorigen. Als Quellen liegen der Präsentation Zeitungen, Photographien, ein gedrucktes Reichsgesetzblatt und Aktenmaterial zugrunde. Neben der analog zu den vorigen Modellen erfolgten Aufbereitung des Quellenmaterials wird auch eine tabellarische Zusammenfassung der aktenkundlichen Merkmale eines Behördenschriftstücks aus dem *Dritten Reich* angeboten.

Da Persönlichkeiten und ihr Handeln bei den dargestellten historischen Vorgängen eine besondere Rolle spielen, gibt es innerhalb dieses Präsentationsmodells nicht nur Themen- und Quellenseiten, sondern auch Seiten, die bestimmten Personen gewidmet sind. Bei den in die Präsentation aufgenommenen fotografischen Portraits handelt es sich zum Teil um Paßbilder, die auf

Tableaus montiert sind. Der Benutzer kann zunächst das Paßbild zusammen mit dem Tableau, also die Archivalie als Ganzes, betrachten, wenn er dann mit dem Mauszeiger ein bestimmtes Symbol berührt, wird das Paßbild sofort ohne seine Umrahmung in Vergrößerung sichtbar.

Ein Orts- und Personenindex sowie Literaturangaben bringen auch hier einen Mehrwert für die Benutzung. Außerdem wurde ein chronologisches Gesamtverzeichnis der verwendeten Quellen erstellt, das mit den Quellen- und Themenseiten über Links verknüpft ist. Darüber hinaus wird innerhalb dieser Präsentation der Originaltext der gedruckten Fassung in seiner Gesamtheit zur Verfügung gestellt, somit kann der Text, etwa zum Nachlesen oder Aufbewahren, als Ganzes ausgedruckt werden. Der Originaltext ist zusätzlich durch ein sogenanntes Blitzregister erschlossen. Sobald sich der Mauspfeil über einem Stichwort im rechten Fenster des Registers befindet (ohne Anklicken), erscheint die zugehörige Textstelle im Hauptfenster.

Modell 4: Augenscheinkarte von 1575

Beim vierten Präsentationsmodell findet mit einem gescannten Makrofiche erstmals eine filmdigitalisierte Quelle Verwendung. Die im Zusammenhang mit einem Rechtsstreit vor dem Reichskammergericht Ende des 16. Jahrhunderts entstandene Augenscheinkarte aus den Beständen des Hauptstaatsarchivs Stuttgart zeigt die Landschaft des Taubergrunds bei Laudenbach und Rinderfeld im heutigen Main-Tauber-Kreis.

Im Mittelpunkt der Präsentation steht die wie ein Landschaftsgemälde gestaltete und als colorierte Federzeichnung ausgeführte Karte mit der dazugehörigen Kartenlegende in Form einer mehrseitigen Akte. Die Karte kann in einer Gesamtansicht in einem eigenem Fenster betrachtet werden. Außerdem steht wieder eine Großansicht mit Erläuterungen zur Verfügung. Diese Großansicht kann von der Kartenauswahlseite aus entweder in das Hauptfenster des Browsers geladen werden oder in ein gesondertes Vollbildfenster, das durch den Wegfall der Menüs und der Statusleiste des Browsers einen größeren Betrachtungsraum bietet.

Auf dem Original der Karte sind in der Großansicht bestimmte Stellen durch Buchstaben oder Zahlen gekennzeichnet, auf die in der Legende Bezug genommen wird. Da diese Eintragungen nur schlecht zu erkennen sind, wurden sie für die Online-Präsentation durch besondere Markierungen deutlicher hervorgehoben bzw. ersetzt. Diese im digitalen Bild nachträglich eingezeichneten Markierungen können jedoch mithilfe der rechts von der Karte befindlichen kreisförmigen Symbole *aus- und eingeschaltet* werden, d.h. das Original kann auch in unbearbeiteter Form betrachtet werden.

Trifft der Mauszeiger nun auf einen der hervorgehobenen Buchstaben oder mit Ziffern versehenen Marksteine, so wird im Anzeigefeld die zugehörige Textstelle aus der Kartenlegende von 1575 angezeigt. Auch die Umschrift der

Ortsnamen und der übrigen handschriftlichen Eintragungen erscheint hier, sobald der Mauspfeil auf die jeweilige Stelle trifft. Durch Anklicken von eingezeichneten Ovalen gelangt der Benutzer zu einer vergrößerten Ansicht des entsprechenden z.B. eine Siedlung oder andere wichtige Details darstellenden Kartenausschnitts. Über einen Link *Vergleichskarten* kann ein Zusatzfenster mit zwei Übersichtskarten geöffnet werden, die die geographische Einordnung der auf der Augenscheinkarte dargestellten Region erleichtern. Für die Kartenlegende, die auf weiteren Seiten im Original präsentiert wird, steht die schon oben erwähnte automatische Transkriptionshilfe zur Verfügung. Literaturangaben vervollständigen die Präsentation.

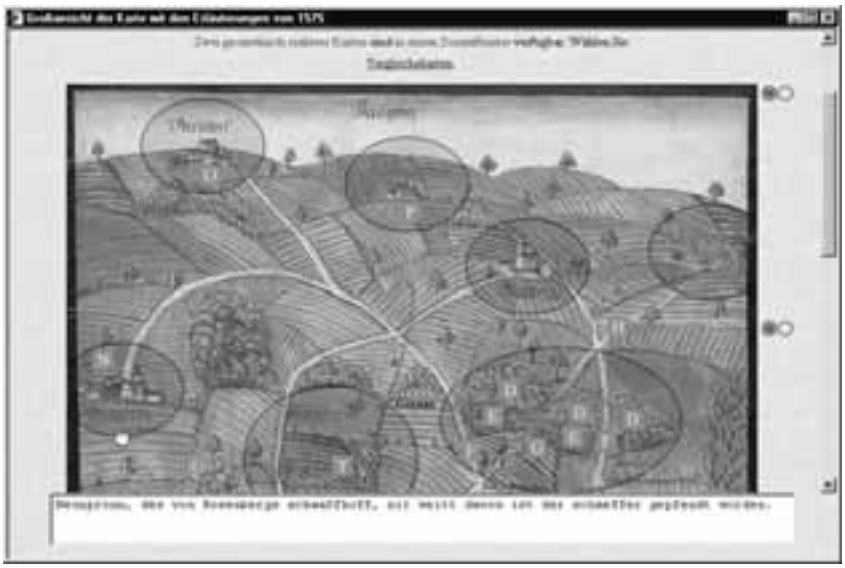


Abb. 3: Internet-Präsentationsmodell *Augenscheinkarte von 1575* (HStA Stuttgart C 3 Bü. 2010). Großansicht mit Markierungen *(http://www.lad-bw.de/digpro/texte/fricke4/augkart.htm).*

Modell 5: Große kurpfälzische Rheinstromkarte von 1595

Das fünfte und aufwendigste Modell, das ebenfalls wieder auf filmdigitalisiertem Material basiert, präsentiert die sogenannte Große kurpfälzische Rheinstromkarte vom Ende des 16. Jahrhunderts. Die im Original über 12 m lange, bildhaft gestaltete Karte zeigt den Rhein von Beinheim (Elsaß) bis unterhalb von Philippsburg. Das Original befindet sich im Generallandesarchiv

Karlsruhe, kann aber wegen seiner ungewöhnlichen Länge und starker Schädigungen dem Benutzer im Archiv nicht mehr vorgelegt werden. Deshalb und wegen ihrer historischen Bedeutung wurde die Rheinstromkarte für die Internetpräsentation ausgewählt.

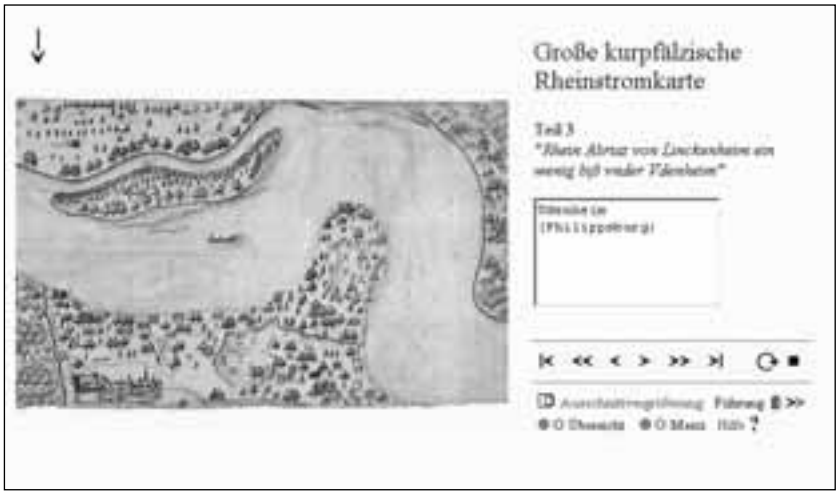


Abb. 4: Internet-Präsentationsmodell *Große kurpfälzische Rheinstromkarte,* Ende 16. Jahrhundert (GLA Karlsruhe, H/Rheinstrom Nr. 19, 24, 27) *(http://www.lad-bw.de/digpro/texte/fricke5/rhein.htm).*

Auch in diesem Modell wurde ein mehrdimensionaler Zugriff auf die einzelnen Kartensegmente realisiert. Über ein besonderes Symbol kann man sich entweder durch die Präsentation führen lassen oder gleich direkt über ein eingeblendetes Auswahlmenü einzelne Seiten anwählen. Außerdem kann standortbezogen direkt von einer Quellenansicht zur anderen übergewechselt werden.

Die Rheinkarte wird in drei Größen gezeigt, zunächst in einer stark verkleinerten, horizontal verschiebbaren Gesamtansicht, wobei zur Erleichterung der Orientierung am Rande die Namen der in der Karte eingezeichneten Ortschaften angegeben sind. Der Verlauf des heutigen Rheins und die heutige Besiedlung des Gebiets kann mit dem früheren Zustand mithilfe zweier moderner Rheinkarten verglichen werden. Eine weitere Ansicht zeigt die Rheinkarte in einer Größe, die noch einen guten Überblick ermöglicht und schon die wesentlichen Details erkennen läßt. Eine Besonderheit ist hier die für eine historische Fluß-Karte objektgerechte Darstellungsart in fortlaufender Form. Die Karte wurde nicht in einzelne quadratische Segmente aufgeteilt, wie dies bei modernen Stadtplänen, die im Internet abrufbar sind, üblich ist. Statt des-

sen kann der Benutzer jeweils einen Teil der im Original dreigeteilten historischen Rheinkarte in seiner gesamten Länge wie auf einer Filmleinwand von links nach rechts Revue passieren lassen. Dabei stehen einem Videorecorder nachempfundene Bedienungselemente zur Verfügung. Durch Anklicken des mit einem Pfeil versehenen kreisförmigen Symbols wird der Vorgang gestartet. Es ist auch möglich, ihn durch Anklicken des quadratischen Pausesymbols zu unterbrechen und durch Anklicken der entsprechenden Pfeilsymbole einen oder mehrere Schritte vor oder zurück zu gehen, sowie direkt zum Anfang oder zum Ende zu springen. Sobald eine Ortschaft unterhalb des senkrecht nach unten gerichteten schwarzen Pfeiles links erscheint, wird im weißen Erläuterungsfeld deren Name sichtbar. Wird die Option *Ausschnittvergrößerung* ausgewählt, so wird sofort zu einer Seite gewechselt, die den gerade unter dem senkrechten Pfeil befindlichen Abschnitt in einer dritten weiter vergrößerten Detailansicht zeigt.

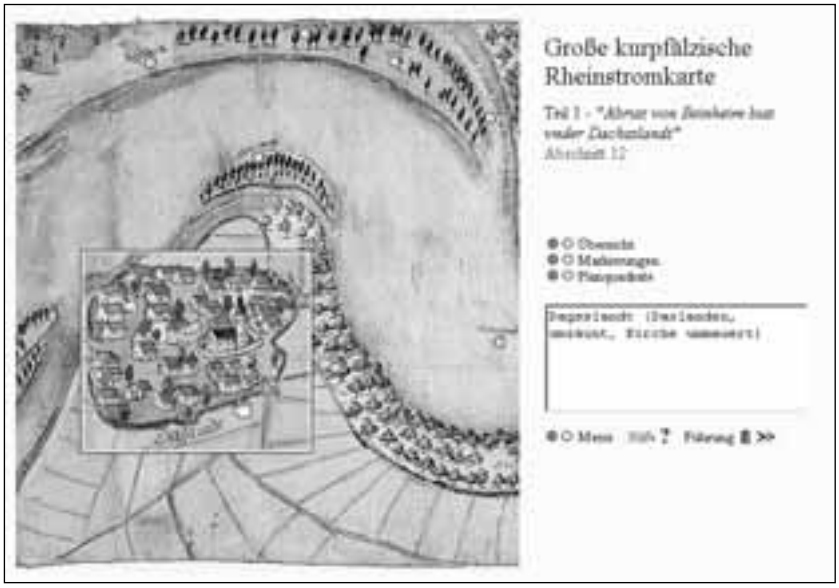


Abb. 5: Internet-Präsentationsmodell *Große kurpfälzische Rheinstromkarte.* Detailansicht.

Die Karte enthält viele handschriftliche Einträge. Um dem Benutzer das Auffinden dieser Einträge auf der Karte zu erleichtern, wurden die entsprechenden Stellen in dieser Detailansicht durch Markierungen in Form gelber Punkte gekennzeichnet. Befindet sich der Mauszeiger über einem Schriftzug, wird die

Transkription der betreffenden Passage in einem Erläuterungsfeld eingeblendet. Besonders interessante zeichnerische Darstellungen (Ortschaften, Menschen, Schiffe usw.) sind durch ein rot-weißes Symbol gekennzeichnet und werden in Vergrößerung angezeigt, wenn der Mauspfeil sich über dem Objekt oder dem Symbol befindet. Sofern es sich um Ortschaften handelt, erscheint im Erläuterungsfeld der transkribierte historische Ortsname zusammen mit dem heutigen Ortsnamen und erklärenden Zusätzen. Die Markierungen können aus- und eingeblendet werden.

Um dem Benutzer auch von diesem Standort, der Datailansicht, aus einen Überblick zu gewähren und das schnelle Überwechseln zu einem anderen Ausschnitt zu ermöglichen, besteht die Möglichkeit eine *Übersicht* anzuwählen. Am oberen Rand des Bildschirms wird dann der jeweilige Teil der Originalkarte in seiner Gesamtheit in der Darstellungsgröße der Gesamtansicht und wiederum mit Angabe der Ortsnamen an den Rändern eingeblendet. Der derzeitige Standort des Benutzers ist durch einen schwarzen Rahmen markiert. Durch Anklicken kann ein beliebiger Kartenabschnitt direkt angewählt werden. Rote Pfeile ermöglichen das Überwechseln von einem zu einem anderen Teil der dreigeteilten Originalkarte.

Auch direkt von dem eingangs erwähnten Auswahlmenü aus kann eine entsprechende Übersichtsdarstellung mit Auswahlmöglichkeit aufgerufen werden, die alle drei Kartenteile umfaßt und zusätzlich Angaben über den Zustand der Karte an einzelnen Stellen enthält. Um die wissenschaftliche Arbeit mit der Karte zu erleichtern, stehen Erläuterungen zu Wappen und historischen Bezeichnungen zur Verfügung, die ebenfalls über das Auswahlmenü erreichbar sind und in einem Zusatzfenster eingeblendet werden. Ein Index der Orts- und Flurnamen mit Referenz auf Planquadrate, die in der Detailansicht der Karte ein- und ausschaltbar sind, sowie ein Sachindex und eine Literaturliste vervollständigen das Modell.

Im Zuge der Entwicklung des Modells wurde auch eine genaue Dokumentation der einzelnen Arbeitsschritte erstellt und die für diese erforderliche Arbeitszeit ermittelt. Während die anderen Modelle unter Verwendung der schon länger eingeführten Programmiersprachen HTML und Javascript erstellt wurden, kam hier auch sogenanntes Dynamic HTML zum Einsatz. Die beiden führenden Browserhersteller Mircrosoft und Netscape haben sich bisher noch nicht auf eine einheitliche Syntax für diese HTML-Erweiterung geeinigt, so daß mit sogenannten Cross-Browser Scripten gearbeitet werden mußte, wodurch der Arbeitsaufwand für die Erstellung der Präsentation stieg. Dies ist also zu berücksichtigen. Die Gesamtarbeitszeit für das Modell betrug neun Wochen (9 x 38 Stunden) oder 342 Stunden. Die einzelnen Arbeitsschritte weisen folgendes Verhältnis auf:

Arbeitsschritte	in %
1. geschichtswissenschaftliche Recherche: Sammeln von Informationen zum Objekt (Bibliographieren/Bücher beschaffen/lesen)	2
2. Transkription der handschriftlichen Quellentexte	3
3. technische Recherche: im Internet und in aktueller Literatur (neue Browserfunktionalitäten, neue technische Möglichkeiten)	5
4. Ausarbeitung eines Konzepts: inhaltliche Gliederung, Design, Funktionalität	2
5. Beschaffung von Zusatzmaterial (für Vergleichskarten, Logos usw.)	1
6. Texterstellung (Quellenbeschreibung, Hilfefunktion)	2
7. Registererstellung (nur für Teil 1 der Rheinstromkarte und die Beschreibungstexte)	5
8. individuelle Bildaufbereitung (Ermittlung des besten Verfahrens durch zahlreiche Tests)	25
9. Erstellung der Icons, Logos, Vergleichskarten und Hintergründe – scannen, bearbeiten, zeichnen	5
10. Programmierung und Anordnung der Inhalte auf dem Bildschirm, sowie Test und Korrektur	50

Als Ergebnis läßt sich festhalten, daß 50% der Arbeitszeit auf die Programmierung fielen, 25% auf die Bildaufbereitung und 25% auf sonstige Arbeiten wie Konzeption, Recherchen und Texterstellung.

Modell 6: Habsburger Einkünfterodel

Darstellungsobjekt des sechsten Modells ist das *Habsburger Einkünfterodel über schwäbische Gegenden*, ein 2,80 m langer Pergament-Rotulus vom Ende des 13. Jahrhunderts aus dem Hauptstaatsarchiv Stuttgart. Auch hier wurden vorrangig vom Mikrofilm digitalisierte Bilder für die Präsentation verwendet. Autor der wissenschaftlichen Texte ist Robert Kretzschmar.

Die objektgerechte Darstellungsart ist auch bei einem Pergament-Rotulus die fortlaufende Form, nur daß die Laufrichtung nicht, wie bei einer Flußkarte, von links nach rechts, sondern von oben nach unten weist. Die Navigation erfolgt wie bei der Rheinstromkarte. Darüber hinaus können mithilfe eines Pulldown-Menüs gezielt einzelne Abschnitte des Quellentextes angewählt werden. Unmittelbar unterhalb des Rahmens, in dem sich die Quelle bewegt, wird die Umschrift eingeblendet. Die Zeilen der Umschrift laufen mit der

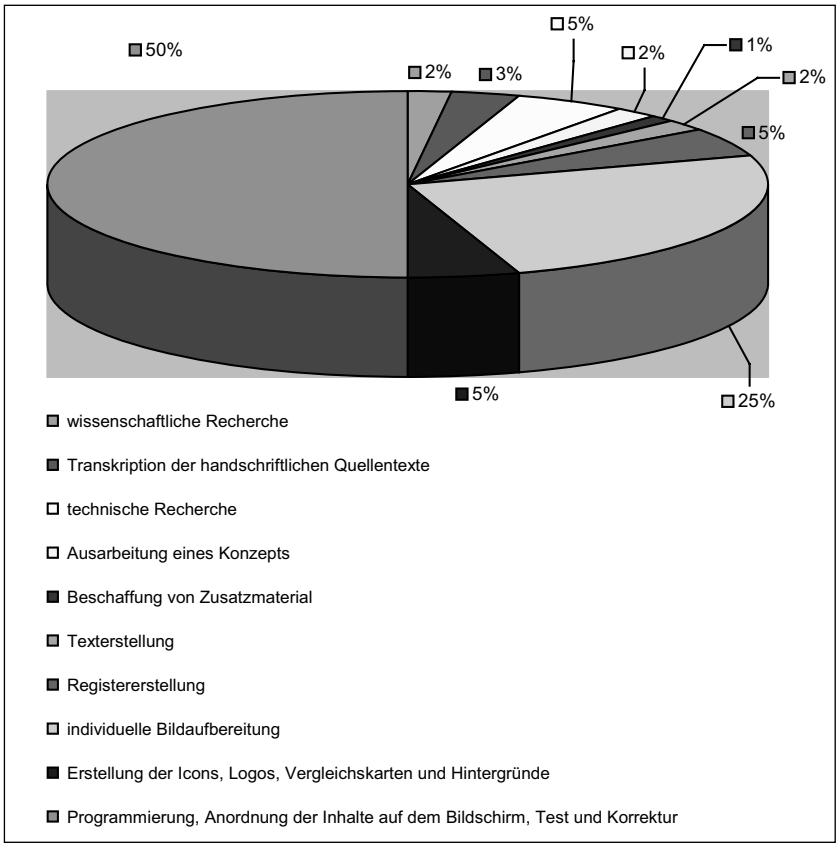


Abb. 6: Präsentationsmodell *Große kurpfälzische Rheinstromkarte von 1595.* Erforderliche Arbeitszeit in %.

Quelle automatisch mit. Anstelle des transkribierten lateinischen Textes kann auch die deutsche Übersetzung eingeblendet werden. Der Benutzer kann frei zwischen Latein und Deutsch hin- und herwechseln, selbst wenn die Durchlauffunktion aktiv ist. Bestimmte erklärungsbedürftige Begriffe im deutschen oder lateinischen Text sind mit Links belegt, bei deren Betätigung sich jeweils ein kleines gesondertes Browserfenster mit Erläuterungen und zum Teil auch mit Abbildungen öffnet. Wie die anderen Modelle enthält auch dieses einen Index mit Verknüpfungen zu den jeweiligen Textstellen und ein Literaturverzeichnis.

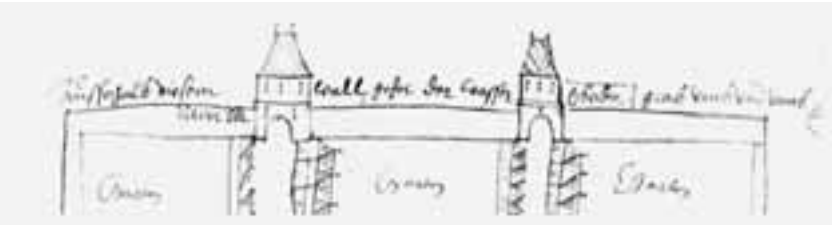
Abb. 7: Internet-Präsentationsmodell *Habsburger Einkünfterodel über schwäbische Gegenden* (HStA Stuttgart H 162 Nr. 1) *(http://www.lad-bw.de/digpro/texte/fricke6).*

Dieses letzte Modell konnte innerhalb der Ausstellung *Vorderösterreich* des Württembergischen Landesmuseums einem größeren Publikum vorgestellt werden. Funktionell durchdachte und optisch ansprechende Internetpräsentationen erfordern zwar einen gewissen Mehraufwand bei der Erstellung, der sich jedoch lohnt, weil das Interesse an der Arbeit mit Archivmaterial und historischen Quellen durch die didaktische Aufbereitung geweckt und gefördert werden kann, sei es nun im schulischen, universitären oder musealen Bereich.

Das Digitale Archiv
Ein Projekt am Stadtarchiv Duderstadt

Von Hans-Heinrich Ebeling

Herzlich willkommen im Stadtarchiv Duderstadt – so empfängt den Besucher die Homepage unseres Projekts seit dem September 1997 mit weit geöffneten Toren einer Stadtansicht (Augenschein) aus dem späten 16. Jahrhundert begrüßt.



Neben Informationen über das Archiv kann der Besucher sich über eine Kurzübersicht der Bestände oder über die Tektonik des Hauses Zugang zu den klassischen Abteilungen – Urkunden, Handschriften, Akten, Sammlungen etc. des Archivs verschaffen, Findbücher anfordern und die Aktentitel über die Klassifikationen oder über Volltextrecherche mit Booleschen Operatoren abfragen. In einem erweiterten Suchformular sind Einschränkungen der Suche nach bestimmten Zeitfiltern oder unter Einbeziehung oder Ausschluß bestimmter Bestände möglich.

Die Datenbankrecherche ist eine Funktion, die allerdings nur im Internet zur Verfügung steht.

Im Folgenden sollen die Aspekte unseres Projektes in seiner Genese vorgestellt werden[1].

---

[1] Hans-Heinrich *Ebeling*, Hans-Reinhard *Fricke*, Peter *Hoheisel*, Malte *Rehbein* und Manfred *Thaller:* Vom digitalen Archiv zur digitalen Edition. Göttingen 1998 (Broschüre und CD); Digitale Archive. Die Erschließung und Digitalisierung des Stadtarchivs Duderstadt. Hg. von Hans-Heinrich *Ebeling* und Manfred *Thaller*. Göttingen 1999. Weitere Informationen erhält man am Stadtarchiv Duderstadt, Christian-Blank-Str.1, 37115 Duderstadt oder durch das Projekt *Das Digitale Archiv*, Max-Planck-Institut für Geschichte, Hermann-Föge-Weg 11, 37073 Göttingen.URL: http:*//www.archive.geschichte.mpg.de/dud-d.htm.*Email: *digitales.archiv@mpi-g.gwdg.de.*

Vorbemerkung

Im Februar 1996 begann am Stadtarchiv Duderstadt unter dem Titel *Digitales Archiv* ein auf drei Jahre befristetes Projekt in Zusammenarbeit mit dem Max-Planck-Institut für Geschichte in Göttingen, das von der Volkswagen-Stiftung im Rahmen des Förderschwerpunkts *Archive als Fundus der Forschung* mit Sachmitteln gefördert wird. Die Projektgruppe besteht aus den beiden Antragstellern Manfred Thaller vom Max-Planck-Institut für Geschichte in Göttingen und mir, Leiter des Stadtarchivs, sowie den wissenschaftlichen Mitarbeitern, Hans-Reinhard Fricke, Stefan Aumann (1996–1998) und Peter Hoheisel (seit 1998).

Das Projekt *Digitales Archiv*

Im Rahmen dieses Projekts wurden bisher ca. 74 000 Abbildungen von Archivalien digitalisiert (Stand Okt. 1998) und durch eine Datenbank erschlossen. Im Mittelpunkt der Digitalisierungskampagne standen die Rechnungsbücher der städtischen Kämmerei, ein Bestand aus der Abteilung der Amtsbücher, der seit 1395 in fast ungestörter Überlieferung erhalten ist und der zu der wertvollsten Überlieferung des Stadtarchivs gehört. Trotz seiner hohen Bedeutung war der Bestand nur höchst unzureichend durch einen Findbehelf, Karteikarten mit Titel und Laufzeit, erschlossen.

Der zeitliche Rahmen der Erfassung der Rechnungsbücher liegt zwischen 1395 und 1650. Der Beginn 1395 ergab sich aus der ältesten erhaltenen Rechnung, das zeitliche Ende der Digitalisierungskampagne wurde auf das Jahr 1650 festgelegt, da bei der Antragstellung im Sommer 1995 nach einer Testphase von einer Jahresleistung von ca. 20 000 Scans ausgegangen wurde. Inhaltlich ließ sich der Einschnitt mit fundamentalen Wandlungen in der Stadtgeschichte nach dem Ende des Dreißigjährigen Krieges begründen.

Da die realisierbare Leistung den geschätzten Wert von 60 000 Scans jedoch weit überschritt, konnten auch alle weiteren Amtsbücher und alle Akten aus dem Stadtarchiv vor 1650 mit einbezogen werden, so daß jetzt kurz vor dem Ende des Projektes von einer kompletten Digitalisierung der Altbestände vor 1650 gesprochen werden kann.

Für das *Digitale Archiv* wurde jedes Archival seitenweise als Bild im Modus mit 256 Graustufen in einer Bildqualität von 300 dpi eingescannt. Es entstehen dabei Bilddateien in einer Auflösung, die auch mit künftigen Bildschirmgrößen (bis 21 Zoll) und Auflösungsparametern sehr benutzerfreundlich angewendet werden können. Die verlustfrei komprimierten Bilddateien werden auf einer CD-ROM gesichert, dabei können etwa 130–150 Bilder im TIF-Format mit durchschnittlichen Größen von ca. 5–6 MB auf einer CD gespeichert werden. Diese Daten werden an zwei Orten – in Göttingen und Du-

derstadt – gelagert und nach der bisherigen Planung durch Umkopieren in einem etwa 10jährigen Zyklus an neue Hard- und Software angepaßt.

Gleichzeitig werden komprimierte Bilddateien hergestellt. Während die sogenannten Archivdateien wegen ihrer hohen Auflösung alle Maßnahmen zur Bildverbesserung – z.B. Lesbarmachung verderbter Stellen, Papieranalysen, Schriftuntersuchungen etc. – ermöglichen, handelt es sich bei den komprimierten Bildern um *Leseexemplare*, die allerdings immer noch eine Vergrößerung um einen Faktor zwei gestatten und deutlich benutzerfreundlicher sind als die Benutzer-Leseschirme für Mikrofilme.

Diese Abbildungen sind hochkomprimiert und sowohl über das Internet zu erreichen als auch auf Anforderung über eine CD (CD-on-demand) zu beziehen. Auf eine solche CD können bis zu 8 000 Bilder im JPEG-Format gespeichert werden. Die Durchschnittsgröße der Dateien beträgt nun noch etwa 100–120 KB.

Das Einscannen erfolgt in der Regel durch einen DIN A 3 Flachbettscanner, da die Archivalien größtenteils im Folio-Format vorliegen und nach den Vorgaben der Rand der Vorlage deutlich erkennbar sein sollte. Auf eine Erfassung in Farbe wurde aus Kostengründen (Speicherkosten) verzichtet, sie erwies sich bei dem vorliegenden Material als verzichtbar. In besonderen Fällen wurde auch mit einem Aufsichtscanner gearbeitet, doch erfordert hier das ständige Nachfocussieren einen erheblichen Mehraufwand an Arbeitszeit.

Die Digitalisierung von Archivalien, die im Stadtarchiv Duderstadt betrieben wird, wird ergänzt durch eine seitenweise Erschließung der Archivalien – jedenfalls der Rechnungsbücher – mit Hilfe einer Datenbank. Da eine befriedigende OCR für mittelalterliche und frühneuzeitliche Schreiberhände bisher nicht vorliegt, wird die inhaltliche Erschließung durch Übertragung in ASCII-Code vorgenommen. Diese Arbeit wird durch einen wissenschaftlichen Mitarbeiter ausgeführt, der über gute paläographische Fähigkeiten wie auch Kenntnisse der städtischen Verwaltungsstrukturen vom 15.–17. Jahrhundert verfügt. Auch am MPI in Göttingen erfolgt neben dem Ausbau der Datenbank eine inhaltliche Bearbeitung.

Die Kämmereirechnungen sind seit dem Ende des 15. Jahrhunderts durch Seitenrubren – z.B. Einnahmen an Schoß (= eine Steuer auf Grundbesitz, Vermögen und Einkünften) und an Zinsen und Renten einerseits, an Ausgaben für den Marstall, den Ratsweinkeller, städtische Boten, Bedienstete, Geschenke für den Stadtherrn usw. andererseits – geordnet.

Die digitalisierten Abbilder dieser Seiten aus den Rechnungsbüchern sind mindestens durch das Seitenrubrum erschlossen. In besonderen Fällen geht die Erschließung jedoch tiefer – bis auf die Ebene der Personennamen. Diese Personennamen werden erschlossen bei allen Steuerzahlern, bei allen Rentenempfängern, bei den städtischen Funktionsträgern, den Bediensteten, den Neubürgern. Der Grund für diesen Entschluß war, daß in einem anderen Projekt des Archivs seit einigen Jahren die Steuerzahler der Stadt für den Aufbau eines Häuserbuchs erfaßt worden sind.

Neben den vom Archiv im Rahmen des Projekts vorgenommenen Erschließungsleistungen kann im Rahmen einer Interaktion von Benutzer und Archiv die Erschließungsleistung eines Benutzers in das System integriert werden. Die Projektgruppe spricht in diesem Fall von einer *dynamischen Erschließung*. Erstes Beispiel ist die Transkription eines Strafbuches von 1530–1545 im Rahmen einer Magisterarbeit, für die die Bearbeiterin eine CD mit ca. 400 digitalisierten Seiten erhielt und als Belegexemplar eine Transkription in elektronischer Form ablieferte. Im Archiv wird im Moment die Verknüpfung dieser transkribierten Texte mit den Rechnungsbuchseiten durchgeführt.

Seit dem Herbst dieses Jahres ist das System der vorliegenden Rubren der städtischen Rechnungen in ein idealtypisches Schema gebracht und durch ein sogenanntes *Codebook* systematisch erschlossen. Durch das Anklicken eines Begriffs aus den Ausgaben oder Einnahmen der Kämmerei und einen Suchbefehl erhält man eine Liste der vorhandenen Archivalien, nach einem weiteren Mausklick ein Bild der Seite, auf der dieser Begriff auftaucht. So erfährt der Benutzer beispielsweise, daß der Begriff Judenschoß erstmals in der ältesten komplett erhaltenen Liste auftaucht und zuletzt 1500 verwendet wird.

In Vorbereitung ist die Integration von etwa 150 000 bereits erfaßten Personennamen ehemaliger Bürger und Steuerzahler. Mit diesen Namen sollen ausgiebige Rechercheaufgaben erfüllt werden können, Erforschung sozialer Karrieren, bestimmter Gruppen und Eliten, eine Sozialtopographie der Stadt erstellt werden usw. Über die Verknüpfung der Namen mit anderen Archivalien sollen bisher unzureichend erschlossene Archivalien für die Forschung besser nutzbar sein. Einer der Anwendungsbereiche ist etwa die Entschlüsselung von Initialen in Inschriften in der Stadt oder über Überresten der materiellen Sachkultur bei handwerklichen Produkten.

Als Benutzungsoberfläche für den normalen Benutzer im Lesesaal oder im Internet entschied sich die Projektgruppe für eine Maske eines Netbrowsers (Netscape oder Internet Explorer), im Hintergrund steht das sehr mächtige Datenbanksystem Klio, das die benötigten HTML Dateien aus den sogenannten Rohdaten in automatisierten Verfahren erzeugt, wenn Anfragen an die Datenbank gerichtet werden.

Bei der ursprünglichen Konzeption des Projekts war vor allem an mögliche neue Zugehensweisen für ungeübte Archivbenutzer durch Verwendung graphischer Oberflächen gedacht worden, z.B. die Möglichkeit über einen Stadtplan Informationen zur Stadtgeschichte abzurufen. Nach ausgiebigen Diskussionen mit Archivaren und Historikern kam die Projektgruppe zu dem Entschluß dazu zurückzukehren, auch den klassischen Weg zur Archivbenutzung digital nachzubilden. Auf dem Bildschirm verfolgt der Interessierte den üblichen Weg in ein Archiv, man hat den virtuellen Zugang zu unserer Organisationseinheit, erfährt Öffnungszeiten und Nutzungsbedingungen, kann sich anhand einer Kurzübersicht einen Überblick über unsere Bestände verschaffen, ein elektronisches Findbuch bestellen und darin entweder blättern oder über Volltext recherchieren. Schutzwürdige und dem Datenschutz unterlie-

gende Aktentitel sind nicht im Internet.

Zwar steht im Mittelpunkt unserer Bemühungen weiterhin der Bestand Amtsbücher, aber man kann auch Aktentitel aus verschiedenen Beständen abfragen oder Urkundenregesten recherchieren und zusätzlich deren digitale Abbilder am Bildschirm einsehen. Die Vorbereitung einer Archivreise durch Vorbestellungen via email ist möglich, in Vorbereitung ist auch ein *Warenkorb* mit Bestellungen der ausgewählten Aktentitel.

Technik und Kosten

Für das Projekt sind zwei Pentium 100–Rechner mit jeweils 32 MB RAM und 1 bzw. 2,1 GB Massenspeicher im Einsatz, an denen die Datenerfassung und die Sicherung auf CD erfolgt. Seit 1997 ist auch ein Rechner im Benutzersaal verfügbar, ein Pentium 200 mit 128 MB RAM und 9 GB Massenspeicher, an den ein CD-Changer mit sechs Laufwerken angeschlossen ist. Wesentlicher Kostenfaktor dürften heute noch die Großbildschirme (17 und 19 Zoll) sowie der A3-Scanner sein. Der Kostenrahmen von ca. 50 000 DM von 1996 dürfte heute bei höherer Leistung der Rechner deutlich unter 20 000 DM liegen, so daß der Personalkostenanteil für das Scannen immer deutlicher ins Gewicht fällt. Er liegt gegenwärtig bei ca. 30 000 DM im Jahr.

Einen besonderen Preisvorteil zeigt das System bei der Bereitstellung von Kopien für Benutzer – danach könnten je nach Zeitaufwand von Personal für die Erstellung der Batch-Datei bis zu 8 000 Kopien ab 25.– DM möglich sein (CD-Rohlings, Abschreibung, Zeitaufwand), ebenso bei Ausdrucken über Laser-Printer in guter Qualität.

Insgesamt arbeitet die Projektgruppe heute mit einer Datenbank von etwa 10 MB und 74 000 Bilddateien von etwa 10 GB, die im Stadtarchiv und auf dem Rechner der GWDG vorgehalten werden.

Vorteile und Nachteile, Ziele und Perspektiven

Die Digitalisierung von Archivalien beinhaltet immer auch den Gedanken der Vervielfältigung durch verlustfreies Kopieren. Für die Archivwelt, die sich bisher u.a. dadurch definierte, daß sie im Gegensatz zu den Bibliotheken immer nur mit dem Unikat, dem unersetzlichen Einzelstück zu tun hatte, muß ein Umdenken einsetzen. Es bricht eine Revolution für die Archive ein, die mit der Erfindung des Buchdrucks vor 500 Jahren für die Bibliotheken vergleichbar ist.

Für den Archivar gibt es ab jetzt den *Benutzer ohne Gesicht* (M. Thaller), den anonymen User im Internet, der mit oder ohne Anmeldung, mit oder ohne Beratung die Archivalien nutzt. Dem Archivbenutzer geht die sinnliche Erfahrung eines Pergamentbandes oder einer kostbaren Handschrift ebenso

verloren wie das Beratungsgespräch.

Andererseits gewinnt das Archiv Benutzer, die sonst nie eine weite und teure Reise zum Archivort unternommen hätten. Die Archivalien werden geschont vor Tageslicht, mehrfacher Bewegung, mechanischer Belastung und Handschweiß. Der Benutzer erhält binnen Sekunden und ohne Wartezeiten ein sehr gut lesbares Abbild des Archivals und kann diese Bilder jederzeit, unabhängig von Warte- oder Öffnungszeiten benutzen. Mit Hilfe des Datenbankzugriffs kann er sehr gezielt genau jene Seiten aufsuchen, die er benötigt, beispielsweise für eine Langzeituntersuchung über städtische Ausgaben für das Bauwesen in einem Zeitraum über 100 oder 200 Jahre.

Der *Benutzer ohne Gesicht* entspricht dem anonymen Mittelalterhistoriker, der auf ein gedrucktes Urkundenbuch zurückgreift und dessen Arbeit dem Archiv ebenfalls nicht bekannt wird. Wie das Urkundenbuch kann das *Digitale Archiv* zwei Dinge tun: Indem es Material in beliebig vervielfältigbarer Form bereitstellt, regt es die Forschung an und sichert einmalige Überlieferung vor der Vernichtung.

Seit der Vorstellung des Projektes im September 1997 bis zu einer ersten Zwischenbilanz im Mai 1998 hat es über 86 000 Zugriffe auf den Server in Göttingen gegeben, vor allem aus den Rechenzentren deutscher Universitäten, auch aus dem benachbarten Ausland, v.a. von dort, wo Deutsch geläufig ist, aber auch aus den USA. Ein besonders ernsthaftes Interesse ist dort zu vermuten, wo über längere Zeiträume Archivalien in der Größenordnung von mehr als 1 000 Abbildungen angefordert wurden.

Die dauerhafte Speicherung der gewonnenen Daten ist angesichts des hohen personellen Einsatzes bei ihrer Gewinnung von entscheidender Bedeutung. Von dem Anbieter Kodak, Hersteller von Mikrofilmen wie von CDs, werden als Garantiezeiten 500 bzw. 110 Jahre genannt.

Eine Zeitspanne von einem Jahrhundert ist für ein Archiv eher ein kurzfristiger Zeitraum. Tatsächlich geht es aber um noch schnellere Zyklen der Datenmigration. Die CD wird in absehbarer Zeit dem DVD-Standard weichen. Wir werden dann umkopieren müssen – aus dem *ewigen Datenträger* (z.B. Papier oder Pergament) wird ein *ewiger Datensatz*.

Das Verfahren bringt Kosten mit sich, die aber deswegen als minimal eingestuft werden können, weil sie in automatisierten Verfahren durchgeführt werden können. Nach dem gegenwärtigen Stand würden durch die Datenmigration für ca. 1 200 CDs in Archivqualität Materialkosten von ca. 3 500 DM entstehen. Dafür erhält man ca. 250 Jahrgänge städtischer Rechnungsbücher verlustfrei kopiert. Die Restaurierung des *ewigen Datenträgers*, eines durch Wasser und Pilzbefall geschädigten Papierbandes eines einzigen Jahrgangs der Rechnungsbücher kostet mindestens ebensoviel.

Allerdings ist die ständige Überwachung und die dauernde Migration der Daten im Zyklus der technischen Innovationen ein zunächst neuer und erschreckender Gedanke.

Bei dem Projekt *Das digitale Archiv* handelt es sich um ein Modellprojekt,

in dem Module entwickelt worden sind, die es auf andere Archive und Projekte übertragbar machen. Die denkbaren Anwendungen sind vielfältig, überall dort wo Material mit Datenbanken geordnet werden kann und wo Abbildungen in digitalisierter Form erzeugt werden können.

Anwendung wird die Form der Digitalisierung von Archivalien zunächst vor allem in Beständen mit vergleichsweise überschaubarem Umfang und hohem Überlieferungswert finden.

Das digitalisierte Buch im Internet
Möglichkeiten der Präsentation und Navigation

Von Norbert Lossau

Vorbemerkung

Die Frage der dem jeweiligen Original gerecht werdenden Digitalisierungstechnik hat die Teilnehmer des Kolloquiums *Digitale Archive und Bibliotheken – Neue Zugangsmöglichkeiten und Nutzungsqualitäten* intensiv beschäftigt und beschäftigt viele bei der konkreten Projektarbeit.

Ein ganz wesentlicher Punkt für die Akzeptanz der von uns digitalisierten Bestände bei Wissenschaftlern und Studenten aber ist die Frage der Bereitstellung von und der Navigation in digitalisierten Dokumenten.

Hierzu sollen im ersten Teil des Beitrags ausgewählte Thesen und Gedanken vorgestellt werden, die dann im zweiten Teil anhand einiger Beispiele aus der Praxis der Online-Präsentation digitalisierter Bücher veranschaulicht werden.

Wozu das Buch digitalisieren?

Die Gründe für eine Digitalisierung von Büchern lassen sich grundlegend nach zwei Aspekten unterscheiden:

- Pragmatische
- Qualitative

Zu dem pragmatischen Aspekt zählt an vorderster Stelle die Verbesserung der Infrastruktur für Informationen und wird deshalb als Hauptargument für die Einführung des neuen Förderprogramms *Retrospektive Digitalisierung von Bibliotheksbeständen* der Deutschen Forschungsgemeinschaft genannt.[1]

Die Bereitstellung von Texten direkt am Arbeitsplatz des Wissenschaftlers, auch außerhalb der Bibliothek und frei von räumlichen und zeitlichen Beschränkungen können unter diesen Punkt ebenso subsummiert werden wie

---

[1] URL: *http://www.dbi-berlin.de/projekte/d_lib/leitproj/vdf/retrodig/retrodig.htm.*

die Möglichkeit des Zugriffs auf Bestände, die in der eigenen Bibliothek nicht vorhanden oder aus konservatorischen Gründen nur eingeschränkt benutzbar sind. Bisher waren hierfür nicht selten zeit- und kostenintensive Reisen direkt an die besitzende Bibliothek erforderlich oder es war der Weg über die Benutzung einer Mikroform zu gehen, deren Vorteile für die Langfristarchivierung zwar unbestritten sind (zumindest für S/W-Materialien), deren Benutzerfreundlichkeit jedoch sehr unterschiedlich beurteilt wird.

Qualitative Aspekte beinhalten eine neue Qualität der Arbeit mit Primärtexten. Erweiterte Arten des Zugriffs auf den Text sowie die parallele Nutzung mehrerer Auflagen sind nur einige Beispiele für diese neue Qualität. Voraussetzung für eine solche Nutzung ist natürlich die Volltextdigitalisierung, d.h., die Umwandlung des durch das Scannen gewonnenen digitalen Faksimiles in einen nach Stichwörtern durchsuchbaren Text. Ist die ökonomische Bearbeitung großer Textkorpora gegeben – was im Falle der Frakturschrift zur Zeit noch große Probleme bereitet – können die Texte mit bekannten und neuen Fragestellungen untersucht werden. Die Frage von Interferenzen zwischen verschiedenen Texten in der Literaturwissenschaft mag ebenso dazugehören wie Fragestellungen in der historischen, deskriptiven Übersetzungswissenschaft, die beispielsweise ein ganzes Korpus von Übersetzungen ein und desselben Werkes in den Mittelpunkt ihrer Untersuchungen stellt.

Benutzung von Büchern: konventionell/digital

Konventionelle Literatursuche

Wie stellt sich die Benutzung von Literatur in wissenschaftlichen Bibliotheken konventionell dar? Die Literatursuche erfolgt im allgemeinen klassisch über den Katalog der Bibliothek, der entweder noch in Zettelform ruht oder bereits elektronisch in einen Online-Katalog konvertiert wurde. Es wird bibliographisch und/oder sachlich auf *Beschreibungsebene* recherchiert, komfortabel ist dabei eine mögliche Kombination von Sucheingabe und Listensuche.

Die moderne Freihand-Bibliothek bietet dazu eine visuelle Benutzung des Bestandes. Durch das Blättern in dem einzelnen Buch wird dokumentspezifisch eine intensivere Suche ermöglicht. Die konventionelle Literatursuche kann so überblicksartig wie folgt charakterisiert werden:

- Katalogsuche (klass. bibliogr. Kategorien): *unbegrenzt* große Literaturmenge (*anonym*), Zugriff auf das Dokument als Ganzes
- Freihand (Regal *browsen*) : Visuelle Suche (begrenzt auf das jeweilige Blickfeld, einprägsam, Zugriff auf Teile des Dokuments möglich, eigene Abstractbildung)

Digitale Literatursuche

Die Zusammenführung der Recherche auf Beschreibungsebene mit der visuellen Suche wird durch das digitale Medium ermöglicht. Ergebnisse aus Erschließungsprojekten (Index deutschsprachiger Zeitschriften/Rezensionsindex) lassen sich so sinnvoll mit Image-Abbildungen der Zeitschrift (d.h. der erschlossenen Quelle) verknüpfen. Der Wunsch nach Visualisierung bei der Literatursuche, wie er beispielsweise auch in DFG-Projekten wie dem VD 17[2] deutlich wird und die Arbeit an den Originalquellen kann durch die Digitalisierung und Online-Bereitstellung befriedigt werden und unter dem Gesichtspunkt der Wissenschaftshermeneutik so eine Vielzahl von Forschungserkenntnissen zutage bringen, die alle an ein- und derselben Originalquelle gewonnen wurden.

Die Möglichkeit der Schaffung von Forschungskorpora durch wissenschaftlich begründete Quellenzusammenstellung sowie eine Verknüpfung mit vorhandener Sekundärliteratur bietet zudem einen besonderen Reiz bei der wissenschaftlichen Arbeit.

Die Erschließung des Buches bis auf Seitenebene herab ist dabei unabdingbar für den komfortablen Zugriff auf die digitale Version.

Modelle der Präsentation des Buches in digitaler Form

Grundsätzlich lassen sich folgende Typen der Präsentation unterscheiden:

1. Nur Images
2. Images + Erschließung
3. Images + Text
4. Nur Text

Zu 1.:
Allenfalls für Einzelbilder oder kurze Bildsequenzen geeignet und wird deshalb gerne für Bildersammlungen, kleinere Handschriften, Einblattdrucke u.a. verwendet.

Zu 2.:
Eine Reihe von Digitalisierungsprojekten in Deutschland widmet sich älteren Drucken vom 17.–19. Jahrhundert. Die Verwendung der Frakturschrift in diesen Drucken verbietet zur Zeit noch eine ökonomisch vertretbare Konversion in maschinenlesbaren Volltext. Auf Empfehlung der AG Inhalt der Deutschen

---

[2] URL: *http://www.vd17.bsb.badw-muenchen.de.*

Forschungsgemeinschaft zur retrospektiven Digitalisierung soll der Nachteil der reinen Imagedigitalisierung von Texten durch ergänzenden Erschließungsaufwand teilweise ausgeglichen werden. Durch die Bereitstellung elektronischer Inhaltsverzeichnisse, Register und anderer durch das Buch bereits vorgegebener Erschließungshilfen wird die elektronische Verknüpfung einen gezielten Zugriff auf einzelne Imageseiten erlauben und so einen wesentlichen Zugewinn an Nutzungsmöglichkeiten bieten.

Zu 3.:
Eine in internationalen Digitalisierungsprojekten wie ILEJ[3], JSTOR[4] anzutreffende Form der Präsentation, die eine Visualisierung des Originals auf dem Bildschirm mit den Möglichkeiten der Volltextrecherche im Hintergrund verknüpft. Treffer werden im elektronischen Faksimile markiert (*gehighlighted*) angezeigt. Neben der Bewahrung des Original-Layouts bei der Online-Präsentation bietet sich für die Volltext-Erfassung auch die Möglichkeit, höhere Fehlerquoten zu tolerieren, als dies bei einer sichtbaren Präsentation des Volltextes möglich wäre.

Zu 4.:
Wird u.a. bei der Digitalisierung von Wörterbüchern verwendet und ist zudem in kommerziell vermarkteten Produkten wie Literatur-CD-ROMs häufig anzutreffen.

Die Schnittstelle zwischen Daten und Internet: Parameter zur Gestaltung des Web-Interfaces

Die Benutzung des digitalisierten Buches im Internet wird durch das *Web-Interface*, der Schnittstelle zwischen digitalen Rohdaten und Web-Browser bestimmt. Technik und Design dieser Schnittstelle, dem *Freihand-Regal* für das digitale Buch, werden durch eine Vielzahl von Faktoren beeinflußt, von denen einige kurz genannt werden sollen und die im zweiten Teil des Vortrags anhand von konkreten Beispielen näher erläutert werden:

Frames

Erlauben eine gute Navigation, können aber Probleme beim Hyperlinking und Ausdruck mit sich führen

---

[3] URL: *http://www.bodley.ox.ac.uk/ilej.*
[4] URL: *http://www.jstor.org.*

Navigationsmenü

Wurden im Buch vorhandene Erschließungsmittel (Inhaltsverzeichnis, Register, Abbildungsverzeichnisse) elektronisch umgesetzt und erlauben sie das direkte Ansteuern einzelner Seiten? Wie navigiere ich innerhalb eines Bandes und wie ist die Anbindung zurück an den Sucheinstieg (z.B. Bibliothekskatalog)? Sind verwendete Icons für Buttons systemübergreifend sprechend?

Aufbereitung von Trefferlisten

Übersichtlich, klar strukturiert vs. lange Listen von Einzelseiten eines Buches

Ladezeiten (Dateigröße – Bilder, Farbe – Netzanbindung, Serverleistung)

Kernfragen der Online-Bereitstellung, insbesondere auch für Graustufen- und Farbbilder

View-Komponente

In einer Reihe von abgeschlossenen Digitalisierungsprojekten (z.B. American Memory) werden dem Benutzer Dateiformate über das World Wide Web angeboten, die von Web-Browsern noch nicht standardmäßig unterstützt werden (TIFF, SGML). Hintergrund für diese Art der Bereitstellung ist eine dezidierte Quellenorientierung der dortigen Projekte. Im Vordergrund steht die Optimierung der Retrievalmöglichkeiten eines in SGML konvertierten Textes, die nicht durch die übliche Volltextsuche abgedeckt werden kann.

Im Förderprogramm der deutschen Forschungsgemeinschaft wird dagegen die Benutzung der digitalisierten Texte mit Standard-Web-Browsern eindeutig empfohlen.

Das Scrollen von Bildschirmseiten, zumal wenn es vertikal und horizontal erforderlich ist, hat schon so manchen Benutzer abgeschreckt und sollte deshalb auf ein Minimum reduziert werden.

Das digitale Buch im Internet

Die Digitalisierung von Bibliotheksmaterialien ist für deutsche Bibliotheken noch relatives Neuland, wenn auch das neue Förderprogramm der Deutschen Forschungsgemeinschaft zur *Retrospektiven Digitalisierung von Bibliotheksbeständen* hier mit seiner Anschubfinanzierung seit 1997 bereits Erstaunliches

geleistet hat. Im internationalen Kontext sind insbesondere die Digitalisierungsbestrebungen in den Vereinigten Staaten, zunehmend aber auch in Großbritannien und Frankreich teilweise weit fortgeschritten. So lief die erste Phase einer großen nationalen Initiative zur digitalen Bereitstellung des kulturellen Erbes in den Vereinigten Staaten, *American Memory*, bereits in den Jahren 1990–1994.

Es verwundert daher nicht, daß die für diesen Beitrag ausgewählten Beispielseiten von Präsentationen digitaler Bücher im Internet aus den USA und Großbritannien kommen.

Das neue Dokumenten-Management-System AGORA

Immerhin ist es aber mittlerweile auch in Deutschland gelungen, an einem der beiden von der Deutschen Forschungsgemeinschaft geförderten überregionalen Digitalisierungszentren, dem GDZ (Göttinger Digitalisierungszentrum) der Niedersächsischen Staats- und Universitätsbibliothek[5], eine technische Lösung für die Bereitstellung digitalisierter Dokumente im Internet zu entwickeln, die es in relativ kurzer Zeit zahlreichen Projekten ermöglichen kann, ihre Materialien weltweit im Internet für die Forschung bereitzustellen.

Das neue Dokumenten-Management-System AGORA[6], dessen gemeinsame Entwicklung durch das GDZ und das Softwarehaus Satz-Rechen-Zentrum in Berlin von der DFG finanziell gefördert wurde, konnte mittlerweile als Prototyp einem großen Interessentenkreis von Bibliotheken und Universitäten vorgestellt werden.[7]

AGORA bietet jetzt:

- Ein Administrationstool, das verschiedene Routinen wie den Import, Export und die Konvertierung von Metadaten und Grafikdateien, sowie die vorlagenorientierte Generierung von semistatischen HTML-Seiten (z.B. Inhaltsverzeichnissen) steuert.
- Ein Konvertierungsmodul zur Generierung von Grafikformaten für die Präsentation im WWW (TIFF, GIF, JPEG, PNG als Ausgangs- und Generierungsformat).

---

5 Das zweite Zentrum befindet sich an der Bayerischen Staatsbibliothek in München

6 URL: *http://www.agora.de.*

7 Am 26. April 1999 fand an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen eine gemeinsame Präsentation von AGORA durch das Göttinger Digitalisierungszentrum (GDZ) und die Firma Satz-Rechen-Zentrum, Berlin, statt.

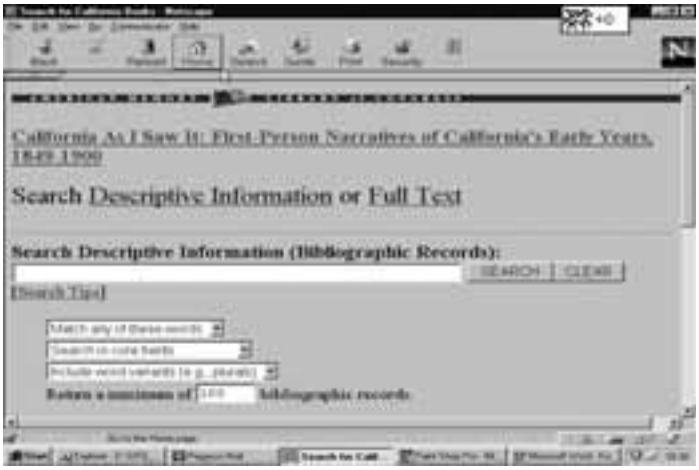
- Für das Retrieval eine *Simple* sowie eine *Advanced Search* über alle Dokumentstrukturen (Autoren, Titel u.a.) der ganzen Bibliothek sowie das Listensuchen (Browsing) über Autoren/Titel. Die dynamische Generierung von HTML-Seiten für Hitlisten, Seitenanzeige aus digitalisierten Büchern, und weitere Navigationshilfen über Java-Servlets. Das Design der HTML-Seiten ist über einfache HTML-Vorlagen projektspezifisch konfigurierbar.
- Die Entkoppelung von Dokumenten- und Web-Server, sowie des Administrationsclients.
- Den Export digitalisierter Dokumente im RDF/XML und strukturiertem PDF-Format für den Download.
- Hardware/Betriebssystem-Plattformen für das Administrationstool: Das Administrationstool ist unter Win95/WinNT4.0 verfügbar. Es greift auf die Metadatenbank über ODBC zu.
- Relationale Datenbank:
  Als SQL-Datenbank wird DB2 von IBM unter SunSolaris auf SUN eingesetzt. Zudem ist DB2 unter WinNT4.0 getestet worden. Generell sind alle SQL-Datenbanken auf den verschiedensten Hardware-Plattformen einsetzbar, die ODBC/JDBC Unterstützung aufweisen.
- Web-Server:
  Als WebServer wird der JavaWebserver von SUN eingesetzt. Verfügbar für SUN Solaris und Win95/WinNT4.0. Einsetzbar sind alle WebServer mit Servletunterstützung (z.B. Apache, Netscape, …) auf den verschiedensten Hardware-Plattformen.

Zusätzlich stellt das GDZ eine eigene Schnittstelle (Java Applikation) für den Import von bibliographischen Daten in das RDF/XML-Format bereit (z.Zt. im PICA 3-Format, geplant ist u.a. auch das Maschinelle Austauschformat für Bibliotheken in der aktuellen Version, MAB 2). Über das http-Protokoll und die Verwendung von CGI-Skripten ist der Zugriff auf die digitalisierten Dokumente über den PICA/GBV-Verbundkatalog möglich.

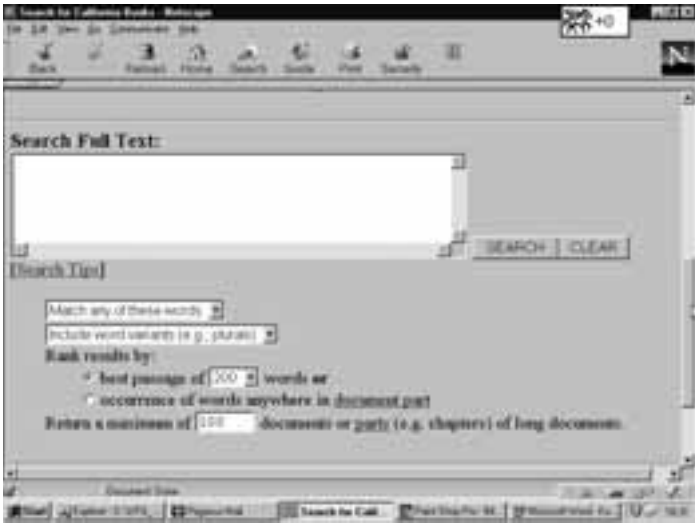
Beispielpräsentationen

American Memory, The Library of Congress[8]

Suche auf Beschreibungsebene (bibliographische Kategorien) und im Volltext



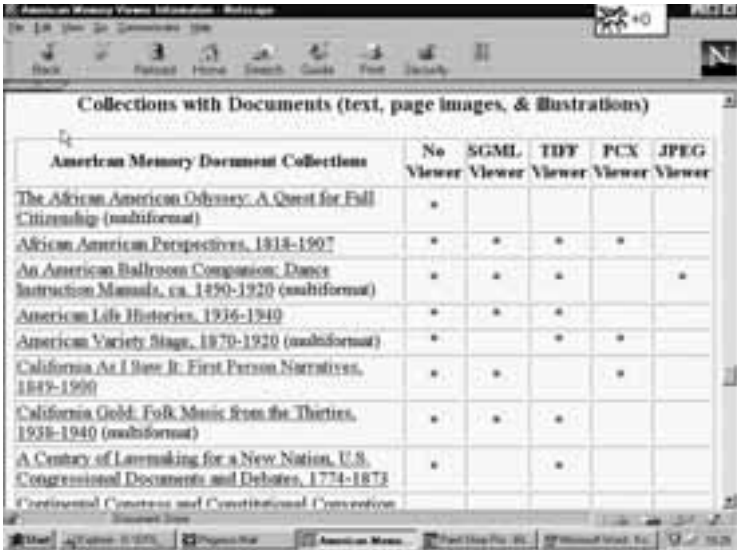
Differenzierte Retrievaloptionen für die Volltextsuche (über mehrere Sammlungen)



8 URL: *http://lcweb2.loc.gov/ammem/ammemhome.html*.

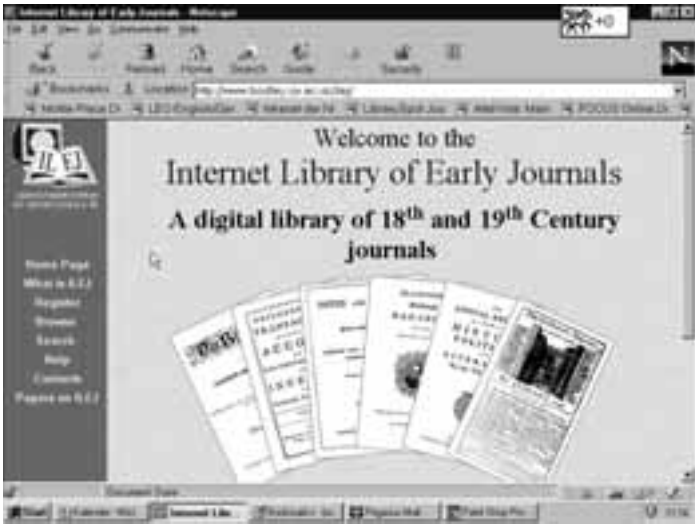
Trefferliste (Vorkommen des Suchausdrucks in Texten aus verschiedenen digitalen Sammlungen)

Notwendigkeit der Verwendung von Plug-Ins für das Anschauen digitaler Dokumente



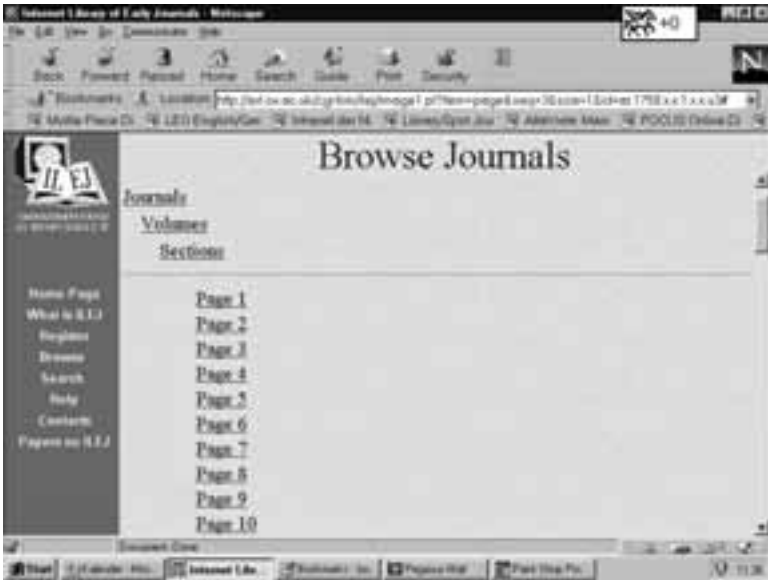
Externes Plug-In für dasAnschauen von SGML-strukturierten Volltexten

Internet Library of Early Journals, An eLib (Electronic Libraries Programme) Project at the Universities of Birmingham, Leeds, Manchester and Oxford[9]

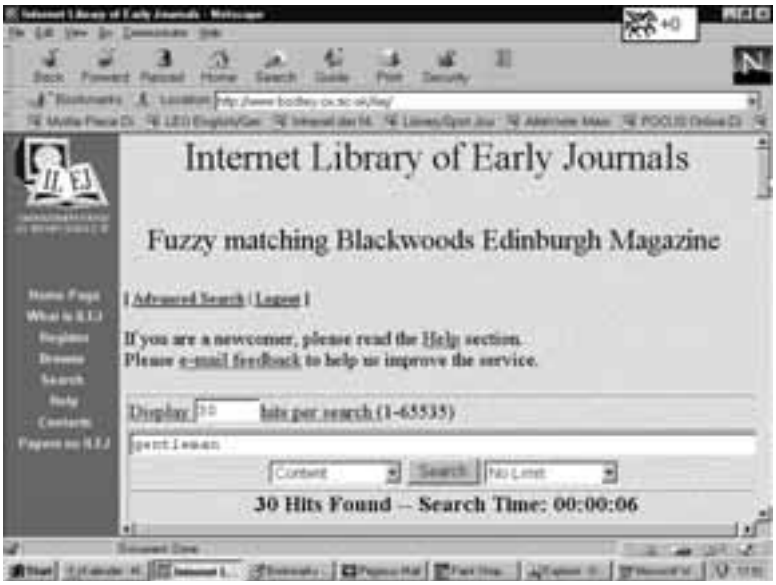


[9] URL: *http://www.bodley.ox.ac.uk/ilej.*

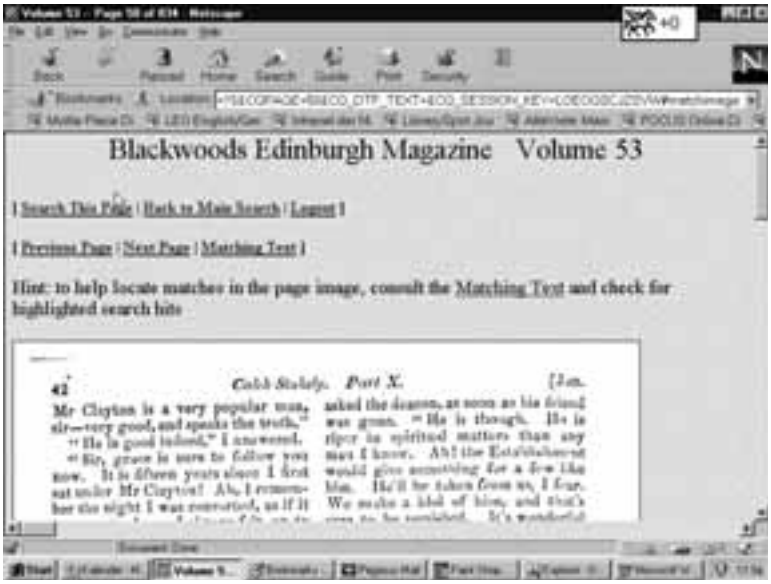
Hierarchische Listensuche über Jahrgang, Band, Artikel



Differenzierte Volltextsuche mit Excalibur (Fuzzy Search u.a.)



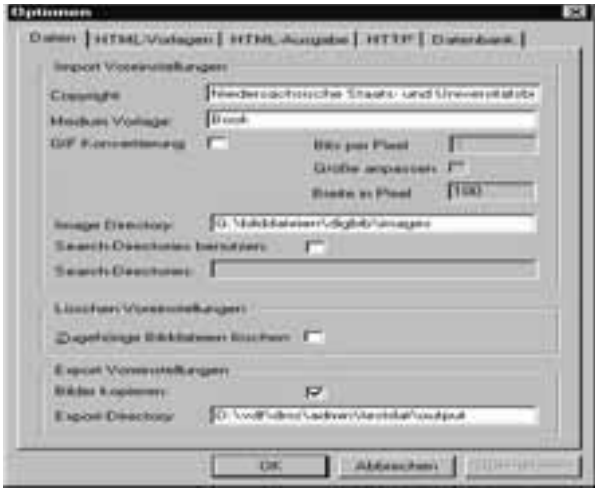
Möglichkeit des Wechsels zwischen elektronischem Faksimile (Image) und (*unsauberem*/raw) ASCII-Text



Administration und Präsentation von Reisebeschreibungen und Nord-Americana mit dem neuen Dokumenten-Management-System AGORA, Niedersächsische Staats- und Universitätsbibliothek/Göttinger Digitalisierungszentrum (GDZ)[10]

Import von digitalen Rohdaten im offenen XML/RDF-Format[11]

Optionenmenü in AGORA (Grafikkonvertierung, Import, Export etc.)



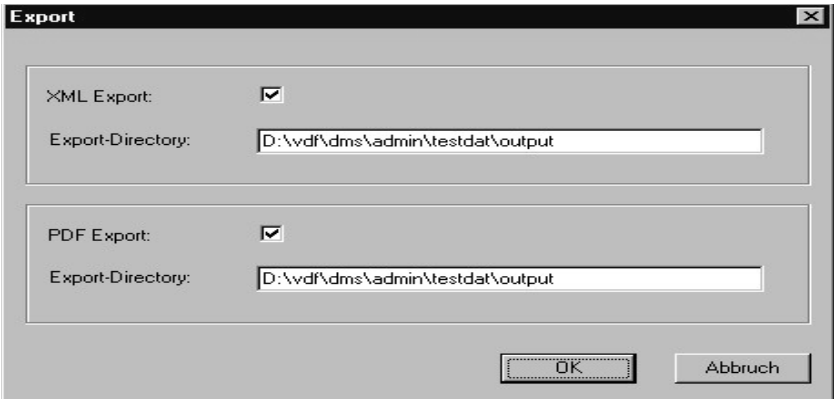
---

10 URL: *http://www.sub.uni-goettingen.de/gdz.*

11 Das Format XML/RDF (eXtensible Markup Language/Resource Description Format) wird vom W3C als Standard für den Austausch von Daten elektronischer Dokumente im Internet empfohlen.

Festlegung von HTML-Templates für die Gestaltung des WWW-Userinterfaces (Zugriff über Java-Servlet)

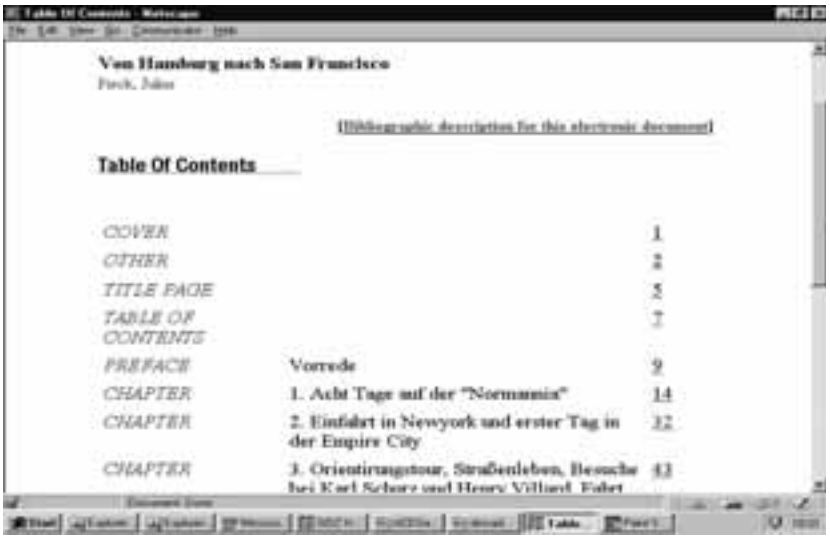
Export der Daten (XML-Format, strukturiertes PDF-Format inkl. Bookmarks z.B. für eine CD-ROM)



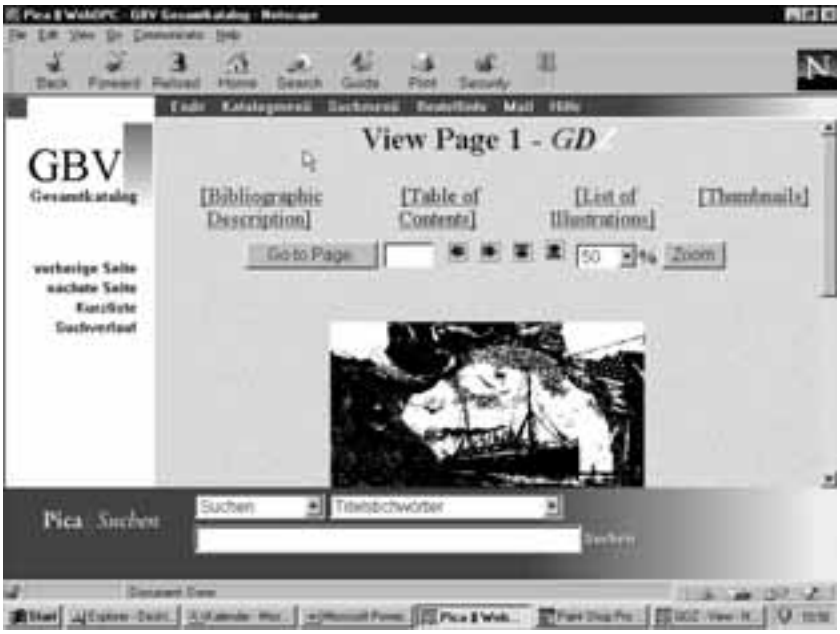
Differenziertes Suchmenü über bibliographische Kategorien, Inhaltsverzeichnisse u.a.

Ansicht einzelner Imageseiten

Elektronisches Inhaltsverzeichnis zum gezielten Ansteuern von Kapiteln etc.



Integrierter Zugriff über den PICA/GBV-Gesamtkatalog



Resümee

National und international liegen mittlerweile eine Reihe von Modellen zur Präsentation von digitalisierten Büchern im Internet vor. Der vorliegende Beitrag hat den theoretischen und praktischen Kontext solcher Präsentationen kurz angerissen und anhand von Beispielpräsentationen aus den USA (American Memory) und GB (ILEJ) illustriert.

Eine neue technische Lösung zur Verwaltung und Präsentation digitaler Bücher im WWW wurde mit dem auf offenen Standards aufbauenden AGORA-System vorgestellt.

Der Mehrwert einer digitalen Präsentation von Büchern läßt sich zusammenfassend an den folgenden Punkten festmachen:

- der pragmatisch erleichterte Zugriff
- der zielgerichtete Zugriff (Erschließungsprojekte)
- der vollständige, vergleichende Zugriff
- der interdisziplinäre Zugriff
- der text- und sammlungsübergreifende Zugriff
- der inhaltliche und visuelle Zugriff

Die Digitalisierungsprojekte im Rahmen des DFG-Förderprogramms zur retrospektiven Digitalisierung sind – unterstützt durch die DFG-geförderten Digitalisierungszentren – auf einem guten Weg, diesen Mehrwert an die *scientific community* zu vermitteln.

Elektronische Bücher im Internet Modell und standardisiertes Werkzeug zur Präsentation sequentiell abfolgender Seiten

Von August Wegmann

Einleitung: Internet und Archive

Das Internet als das weltweite *Informationsarchiv* wird seit langem auch von Archivaren und Bibliothekaren genutzt und natürlich fragen auch normale Endbenutzer nach Möglichkeiten dieses Medium zum einfachen Zugriff auf Archiv- und Bibliotheksgut einzusetzen.

Leider sind zur Zeit die Hürden, Informationen dieser Art im Internet anzubieten, noch immer recht hoch. Ein ganzes Netzwerk von Autoren – Grafiker, Designer, Layouter, Programmierer, Texter u.v.a.m. – wird gebraucht, um die gewünschten Materialien aufzubereiten und als Website zu veröffentlichen.

Deshalb haben wir uns entschlossen, die heute noch anfallende Komplexität für die Präsentation von Archiv- und Bibliotheksgut zu verbergen, den Vorgang der Erstellung einer Website für bestimmte Aufgaben in einer Standardumgebung musterhaft zu lösen und diese Lösung dann als Vorlage – und offenes Werkzeug – für Archivare und Bibliothekare zur Verfügung zu stellen.

Damit sollten Archivare und Bibliothekare eine leichte Möglichkeit zur Publizierung von Materialien erhalten und gleichzeitig alle interessierten Benutzer innerhalb und außerhalb von Archiven und Bibliotheken einen gleichermaßen funktional reichhaltigen wie leicht zu navigierenden Zugang zu diesen Informationen zur Verfügung gestellt bekommen.

Aufgabenstellung: Modell und Werkzeug

Die Aufgabe besteht darin, ein Modell und standardisiertes Werkzeug für eine Präsentation sequentiell abfolgender, d.h. als Images vorliegender Seiten (Buch bzw. mehrseitige Akten) im Internet auf der Basis von HTML und standardisierter Scriptsprachen (u.a. JavaScript, DHTML) ohne Verwendung proprietärer Standards und einer Datenbanktechnologie zu entwickeln.

Unter der Voraussetzung, daß die Erschließungsleistungen (u.a. Inhaltsverzeichnis, Indizes, Titelaufnahmen) schon erbracht sind, soll das Modell einen mehrdimensionalen Zugriff auf das Objekt bieten und folgende Anforderungen sollen dabei berücksichtigt werden:

– sequentielles Blättern über Videorecorder-Knöpfe
– Inhaltsverzeichnis mit Links
– Indizes mit Links
– Historie, Lesezeichen mit Notizfunktion

Als geeignetesTestobjekt wird folgendes vom Mikrofilm digitalisiertes Buch aus dem *großen Warenkorb* des Projekts *Digitale Konversionsformen* der Landesarchivdirektion Baden-Württemberg bereitgestellt: *Krauß, Rudolf: Das Stuttgarter Hoftheater von den ältesten Zeiten bis zur Gegenwart, Stuttgart 1908.*

Lösungsansatz: Erweiterung von Autorensystemen

Moderne Autorensysteme für die Erstellung und Pflege von Webseiten erlauben den Einbau von benutzerspezifischen Erweiterungen. Damit fiel der Entschluß, nicht eine spezielle Software bzw. eine neue Anwendung bzw. einzelne Teile als Java-Applets zu entwickeln, sondern die gesamte Lösung in ein Standardautorensystem einzubetten und in einer reinen HTML-Umgebung mit Einsatz von Skriptsprachen zu realisieren.

Zugleich lag es auch nahe, dem wachsenden Datenvolumen in Bezug auf Texte und Bilder Rechnung zu tragen, und nicht für jedes einzelne Bild eine eigene HTML-Datei als Träger zu erstellen, sondern alle Bilder auf einer einzelnen HTML-Seite nacheinander bzw. auf Benutzeranforderung darzustellen.

Es wurde also darauf verzichtet, für jedes einzelne Bild eine eigene Webseite zu erstellen, was den Pflegeaufwand mit Standardwerkzeugen unnötig verkompliziert hätte. Vielmehr wurde eine abstrakte Lösung vorgezogen, welche innerhalb einer Benutzerseite, die einzelnen Bilder in unterschiedlichen Formaten und Größen darstellt.

Lösung: *Ebuch* als Website-Vorlage

Beim Einsatz von Microsoft Frontpage 98 ist es möglich, eine Vorlage für ein gesamtes Web einschließlich der Navigation, bzw. Vorlagen für einzelne Webseiten zu erstellen. Ein Autor im Archiv- oder Bibliotheksbereich kann eine solche Website-Vorlage für seine eigenen Archivalien und Bücher, welche eine bestimmte Aufbereitung benötigen, benutzen. Nach Einsetzen der spezifischen Parameter bzw. der Anpassung von Standardwerten wird dann aus solch einer Website-Vorlage ein gebrauchsfertiges Web, was auch noch nach eigenen Vorstellungen erweitert und ergänzt werden kann.

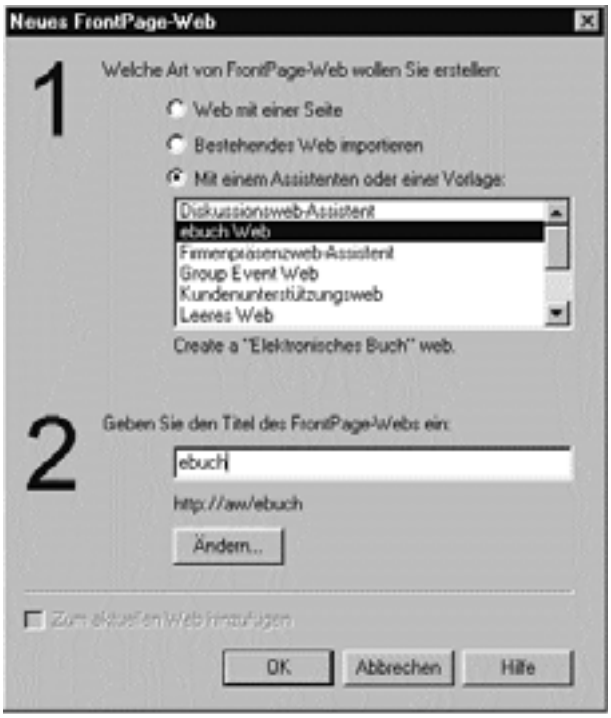
Diese *Ebuch*-Vorlage beinhaltet damit alle für den Betrieb notwendigen Seiten (einschließlich der Navigation), der wünschenswerten Seiten (Benutzerkontakt, Rückmeldung und Diskussionsforum) und stellt einen Rahmen

dar, in dem weitere Seiten in standardisierter Weise eingebunden werden können. Darüber hinaus kann das Aussehen des Web zu einem späteren Zeitpunkt über das *Design*, welches die Darstellung der Standardelemente wie Texte, Listen, Tabellen, Navigationsknöpfe und des Hintergrundes festlegt, angepaßt werden.

Dieser Lösungsansatz erlaubt es dem normalen Mitarbeiter im Archiv, durch Nutzung einer Vorlage mit einem Standardwerkzeug zur Dokumentenbearbeitung (HTML ist Standard) unter Hinzunahme von bereits vorliegenden gescannten Bildern eigene *elektronische Bücher* zu erstellen und im Web zu publizieren. Die Standardanforderungen an die Güte von Benutzeroberflächen im Web in Hinblick auf Klarheit, Konsistenz, Orientierung und Navigation innerhalb der Webseiten wird dabei von der Vorlage erfüllt.

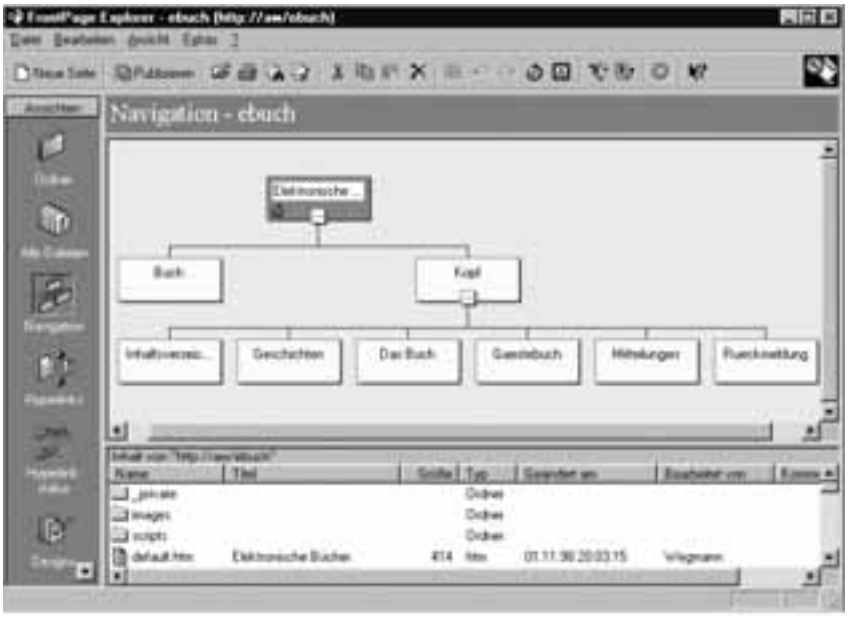
Erstellung eines Buches

Nach dem Start des Frontpage-Explorers kann eine neues Web erstellt werden. Dabei wird standardmäßig eine Dialogbox mit allen verfügbaren Website-Vorlagen angezeigt und der Autor kann eine gewünschte Vorlage auswählen:



Bearbeitung eines Buches

Nach Aktivierung der Ebuch-Vorlage wird selbständig ein – vorerst – leeres elektronisches Buch mit allen notwendigen Seiten erstellt und kann danach vom Autor weiter bearbeitet werden. Es werden die einzelnen Seiten mit den notwendigen Skripten, die dazu passende Navigation, die Ordner für die Bilder und auch eine Liste mit noch auszuführenden Aufgaben erstellt.

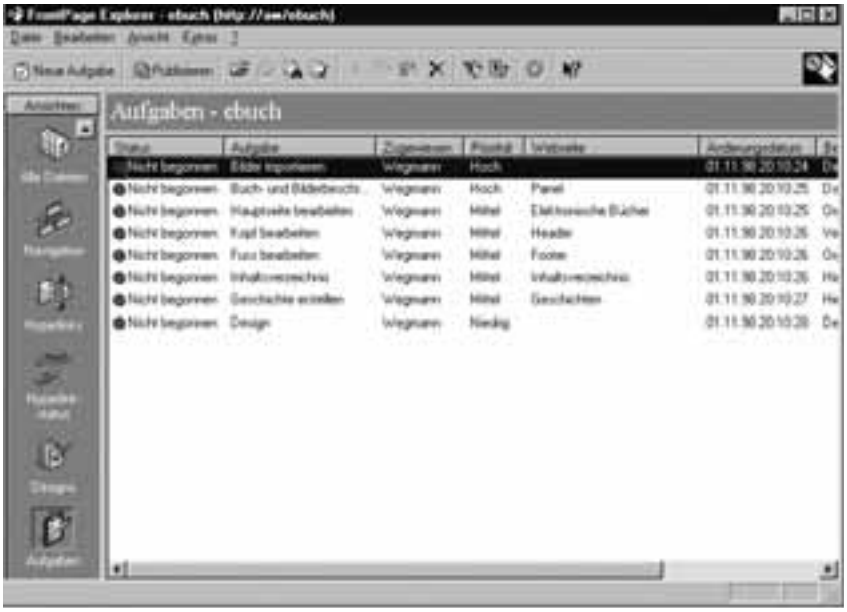


Innerhalb des Frontpage-Explorers werden alle notwendigen Strukturen für ein normales Web erstellt. Damit kann ein solches Buch später genauso wie alle anderen Frontpage-Webs bearbeitet werden und die gesamte Funktionalität der Autorensoftware steht direkt zur Verfügung, ohne daß ein eigenes Autorensystem erstellt wurde. Zusätzlich steht natürlich die Vorlage für das elektronische Buch, das *Ebuch*-Template als Quelltext zur Verfügung und kann auf Wunsch oder Bedarf auch direkt auf äußere Anforderungen angepaßt bzw. erweitert werden, ohne daß Spezialknowhow gebraucht wird. Im Normalfall kann eine Präsentation für entsprechend vorbereitete Materialien sehr schnell erstellt werden. Die eigentliche Pflege kann sich dann auf redaktionelle Texte, Erweiterungen und neues Bildmaterial erstrecken. Eine eigentliche Programmierung wird bei der Erstellung solcher Bücher vor Ort nicht gebraucht, als tiefste Ebene wird zur Zeit die Komplettierung einer HTML-Form mit den eigentlichen Bildinformationen benötigt.

Zu der Standardfunktionalität eines Buches gehören die folgenden Elemente:

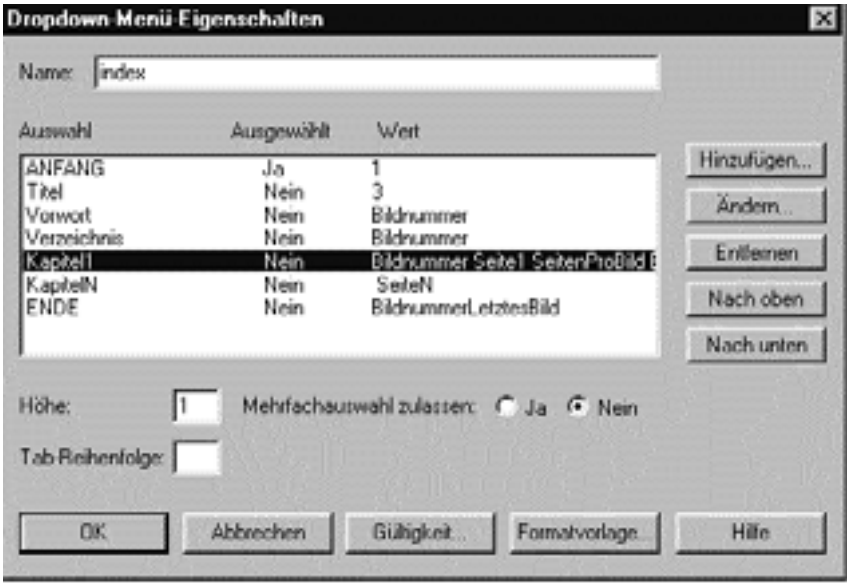
- Die Titelseite des Buches
- Das Inhaltsverzeichnis mit den einzelnen Kapiteln
- *Geschichten* rund um das Buch mit Querverweisen
- Das eigentliche Buch mit
- Objektsichten und Blätterfunktionen,
- Markierung und Anmerkung
- Darstellungsoptionen und Bildformatauswahl
- Das Gästebuch für Eintragungen der Besucher
- Eine Kontaktseite für Mitteilungen und Rückmeldungen

Die noch ausstehenden Aufgaben werden ebenfalls aufgelistet und können so vom Autor einfach sequentiell abgearbeitet werden :



Konfiguration eines Buches

Die Konfiguration eines Buches wird durch Eintragen der entsprechenden Werte innerhalb einer Form realisiert. Diese Form wird dann zur Laufzeit über ein Javascript interpretiert und kann so die Benutzereingaben funktional umsetzen. Dabei werden das Inhaltsverzeichnis, die Objektsichten, Darstellungsgrößen in % und Bildformate bzw. Darstellungsgüten erfaßt.



Beispiel des zugehörigen HTML-Textes:

```
<pa>
<form name="book">
<!— … —>
  <select name="index" size="1">
    <option selected value="1">ANFANG</option>
    <option value="8">Inhaltsverzeichnis</option>
    <option value="10 1 1">Kapitel 1</option>
    <option value="  39">Kapitel 2</option>
    <option value="360">ENDE</option>
  </select>
  <select name="display" size="1">
    <option value="0.1">10%</option>
    <option selected value="0.5">50%</option>
    <option value="1.0">100%</option>
  </select>
  <select name="source" size="1">
    <option selected value="high jpg 741*951">500kb</option>
  </select>
<!— … —>
</form>
<pe>
```

Bearbeitung einzelner Seiten

Standardmäßig werden Seiten im Frontpage-Editor bearbeitet. Dabei können Platzhalter (Kommentare) definiert werden, welche vom Benutzer einfach nur als Handlungsanweisungen gelesen werden bzw. auch textuell überschrieben werden können.

Damit kann der eigentliche Anpassungsvorgang für den Autor transparent gestaltet werden.

Hauptseite eines elektronischen Buches

Nach der Eingangsseite mit einem Titelbild kommt der Benutzer auf die Hauptseite des elektronischen Buches. Diese Hauptseite ist in Frames aufgeteilt und enthält im oberen Bereich eine sogenannte Kopfseite mit dem Titel der publizierten Archivalie und darunter als Linkliste die möglichen Funktionen, im mittleren Bereich wird die jeweils ausgewählte Funktion dargestellt und im unteren Bereich, dem Fußbereich, können allgemeine Informationen zum publizierenden Institut eingeblendet werden.

Diese Aufteilung mit Frames ist für den Endbenutzer gut verständlich und reduziert die Menge der übertragenen Daten auf das notwendige. In der abgebildeten Form wurde auf eine Ausgestaltung der einzelnen Seiten verzichtet;

es ist jedoch möglich, die einzelnen Seiten in diesem Frameset mit weiterem Text oder Bildern bzw. mit Hintergründen auszustatten. Desgleichen können natürlich auch Links auf verwandte Publikationen bzw. Bannerwerbung hinzugefügt werden.

Der Benutzer hat die Möglichkeit – im mittleren Bereich innerhalb des Anzeigeframes – die jeweils aktivierte Funktion zu nutzen. In der Abbildung war das Inhaltsverzeichnis vorgewählt, so daß danach die Seite mit dem Verzeichnis aller Buchinhalte angezeigt wird, und der Benutzer den jeweiligen Text lesen kann und bei Bedarf direkt von dort in das Buch einsteigen kann, indem dem vorgegebenen Link gefolgt wird.

Eine Alternative zum Inhaltsverzeichnis bietet der Menüpunkt *Geschichten*, welcher ebenfalls zum Buch führt. In dieser Auswahl wird dann neben dem Buch in einem weiteren – zum Buch parallelen – Frame eine Seite angezeigt, die Informationen zu allen Buchseiten enthält und die dann jeweils zu der gerade angezeigten Buchseite synchronisiert wird. Damit kann Hintergrundinformation zu einem Buch bereitgestellt werden, welche vom Benutzer direkt recherchierbar ist und auch eine Navigation zu den Buchseiten enthält bzw. je nach ausgewählter Buchseite den dazu passenden Ausschnitt aus den Geschichten zum Buch in den Blickfeld des Betrachters bringt.

Eine Seite im elektronischen Buch:

Das Zentrum des elektronischen Buches bildet die eigentliche Buchseite, welche über die Menüpunkte Inhaltsverzeichnis, Geschichten oder Buch erreicht werden kann. In dieser Buchseite stehen nun alle Funktionen zur Navigation innerhalb des Buches und zur Steuerung der Darstellung der einzelnen Buchseiten zur Verfügung.

Der Benutzer kann sich schnell im Buch bewegen, indem er einen entsprechenden Bereich aus dem Inhaltsverzeichnis auswählt, eine Seite aus dem Buch oder ein Blatt aus den Materialien vornimmt oder einfach über das Videorecorder-Interface mit den Knöpfen vor und zurück die einzelnen Seiten durchblättert. Weiterhin hat er die Möglichkeit, die Größe der gewünschten Auflösung in % zu wählen bzw. in einer Auswahlliste das für ihn passende Bildformat zu selektieren. Somit können dem Benutzer verschiedene Bildformate mit unterschiedlichen Bildgüten zur Verfügung gestellt werden und der Betrachter kann, je nach Arbeitsweise, mit einem niedrigen Bildformat – sprich schneller Übertragungsleistung – beginnen und bei einer genaueren Inspektion dann zu einem höheren Bildformat wechseln, was dann eine längere Übertragungszeit beansprucht. Es ist auch möglich Anmerkungen zu einzelnen Seiten zu machen. Diese werden dann als lokale Cookies gespeichert und bei erneutem Zutritt zur Seite eingeblendet.

Zusätzlich wird natürlich mit dem „?“ eine Hilfe angeboten, welche die einzelnen Interaktionselemente beschreibt und so dem Benutzer erschließt. Für den Betrieb eines elektronischen Buches in einem Vorführraum ist auch eine automatische Blättermaschine eingebaut, welche die einzelnen Bereiche eines Buches automatisch durchläuft und bei einer Eingabe durch einen Besucher den Vorführmodus abbricht und so auch eine normale Nutzung des elektronischen Buches zuläßt.

Ausblick

Das Konzept des elektronischen Buches hat sich in der Praxis bewährt und Archivare oder Bibliothekare können, nach Vorbereitung der sequentiellen Images, daraus innerhalb eines oder weniger Tage publizierbare Websites erstellen. Die Einbettung in ein Standardwerkzeug erlaubt es darüber hinaus die Website nach weiteren Erfordernissen ohne jegliche Einschränkung mit allen verfügbaren Funktionalitäten auszugestalten.

In einer weiteren Planung ist daran gedacht, die Vorlage des elektronischen Buches in einen Assistenten umzuwandeln. Dieser Assistent läuft dann als eigenständiges Programm innerhalb einer Standardumgebung ab und erfragt über eine Anzahl von Dialogschritten die einzelnen Daten zur Erstellung eines Buches ab, um in einen weiteren Schritt die Generierung der gewünschten Website vorzunehmen.

Internet-Präsentationsmodell für Zeitungen und Großformate

Von Thomas Fricke

Eine besondere Herausforderung stellt die Präsentation großformatiger Objekte wie Zeitungen im Internet dar. Hier ist es besonders schwierig eine Lösung zu entwickeln, die trotz der großen Formate einen raschen Datentransfer der einzelnen Dokumente ermöglicht und gleichzeitig eine lesbare Qualität der Dokumente bietet. Am Beispiel der Wochenzeitung *Staatsanzeiger für Baden-Württemberg*, Nr. 1 vom 11.1.1999, 48. Jahrgang, wurde im Rahmen des Projekts *Neue Möglichkeiten und Qualitäten der Zugänglichkeit zu digitalen Konversionsformen gefährdeter Bücher und Archivalien* ein Modell entwickelt, das die genannte Problematik berücksichtigt (siehe *http://www.lad-bw.de/digpro/texte/zeitung*). Das Modell wurde für die aktuelle Browsergeneration, Version 4 des Internet Explorers und Netscape Communicators, konzipiert und mit HTML, Dynamic HTML und Javascript erstellt. Die Funktionalität des Modells soll im folgenden in ihren Grundzügen beschrieben werden.

Die oberste Ebene bildet eine Bestandsübersicht. Hat der Benutzer nun eine bestimmte Ausgabe aus dem Online-Zeitungsbestand ausgewählt, kann er entweder direkt zu einer beliebigen Seite überwechseln oder, um sich einen Überblick über den Inhalt der Zeitung zu verschaffen und nach bestimmten Themen zu suchen, zu einer Liste sämtlicher Artikel der Ausgabe.

Hat der Benutzer sich für die erste der beiden Möglichkeiten entschieden und eine Seite angewählt, so gelangt er zu einer Vorschau, in der er Seitenbereiche, die ihn interessieren, durch Mausklick markieren kann. Die markierten Segmente erhalten dann eine rote Umrahmung. Löst er nun den Link *Ausschnitte betrachten* aus, so öffnet sich ein gesondertes, frei bewegliches Browserfenster, in dem zunächst nur die vorangewählten Seitensegmente geladen und damit sichtbar sind. Durch Scrollen mit den Pfeiltasten kann der Benutzer den ausgewählten Artikel ins Zentrum rücken. Ist das Fenster maximiert steht der gesamte Bildschirm für die Darstellung des Zeitungsausschnitts zur Verfügung, da die Menüs des Browsers ausgeblendet sind. Durch Anklicken der weißen Bereiche im Zusatzfenster können weitere Seitensegmente nachgeladen werden. Es ist auch möglich, ganze Reihen von Segmenten anzuwählen. Die in der Vorschau gesetzten Markierungen können im Bedarfsfall rückgängig gemacht und neu gesetzt werden. Möchte der Benutzer gleich die ganze Seite in lesbarer Größe betrachten, wählt er *ganze Seite*.

Abb. 1: Internetpräsentationsmodell *Staatsanzeiger für Baden-Württemberg (http://www.lad-bw.de/digpro/texte/zeitung).*

Von der Bestandsübersicht aus kann der Benutzer, wie eingangs bereits erwähnt, auch zu einer Artikelliste überwechseln. Diese enthält Titel und Untertitel sämtlicher in der betreffenden Zeitungsausgabe abgedruckten Beiträge. Durch Anklicken der Artikelnummer kann auch von hier aus das Fenster mit der lesbaren Seitenansicht aufgerufen werden, wobei zunächst nur ein Seitensegment vorgeladen wird, um den Beginn des Artikels zu markieren. Weitere Segmente können durch Mausklick, wie beschrieben, nachgeladen werden. Eine Suchfunktion ermöglicht das schnelle Auffinden von Artikeln zu einer bestimmten Thematik.

Das beschriebene technische Konzept der Seitensegmentierung bietet mehrere Vorteile:

- Durch die *Kachelung* der lesbaren Ansicht, also die Aufteilung der Gesamtbilddatei in einzelne Segmente, wird der Ladeprozeß beschleunigt.
- Dadurch daß bestimmte Seitenausschnitte vorangewählt werden können und nur diese angezeigt werden, verringert sich die zu ladende Datenmenge ganz erheblich.
- Das Vorschaubild vermittelt sofort einen Überblick über den Inhalt der Seite, die in lesbarer Größe aufgrund des großen Formats von Zeitungen immer nur in Teilen auf dem Bildschirm angezeigt werden kann.

Abb. 2: Internetpräsentationsmodell *Staatsanzeiger für Baden-Württemberg.* Seitengesamtansicht mit der Möglichkeit weitere Segmente nachzuladen.

Selbstverständlich wäre das Aufteilen jeder digitalisierten Seite in 40 einzelne Quadrate und das zusätzliche Anfertigen eines Vorschaubildes zu jeder Seite sehr arbeitsaufwendig, wenn es unter Verwendung der normalen Funktionalität eines Bildbearbeitungsprogrammes durchgeführt würde, zumal größere Mengen digitalisierter Zeitungen für die Präsentation im Internet aufbereitet werden sollen. Deshalb mußte ein Verfahren entwickelt werden, das eine rationelle Mengenverarbeitung ermöglicht.

Wenn die Zeitungsseiten ein einheitliches Format aufweisen und innerhalb der unbearbeiteten Bilddatei stets gerade und exakt an derselben Stelle positioniert sind, kann die Erstellung vollständig selbsttätig erfolgen. Sämtliche Bilddateien, die Segmente der lesbaren Ansicht und das Vorschaubild, wurden mithilfe einer Photoshop-*Aktion* im Batchbetrieb erstellt.[1] Die zunächst im

[1] Das Zeitungsoriginal ist zunächst schwarzweiß verfilmt und von einem Scandienst bitonal mit einer Auflösung von 400 dpi digitalisiert worden. Da die Seiten nicht gleichmäßig und gerade positioniert waren, mußten pro Seite zwei manuelle Eingriffe vorgenommen werden.

Photoshop-Format abgelegten, aufbereiteten Digitalisate wurden anschließend mit einer anderen Bildverwaltungssoftware, mit der höhere Komprimierungsraten erzielt werden können als mit Photoshop, ebenfalls im Batchverfahren in das PNG-Format konvertiert.

Das Modell soll in einem Folgeprojekt um eine leistungsfähige Volltextsuche erweitert werden. Der Benutzer kann dann innerhalb einzelner Artikel, einzelner Seiten einer Ausgabe oder über mehrere Ausgaben hinweg suchen. Das Modell soll für eine Massenverarbeitung weiter optimiert werden, vor allem was die rationelle Einbindung von Meta- und Bilddaten betrifft.

Digitalisierung im Spannungsfeld von Erhaltung und Zugänglichkeit

Bestandserhaltung in einer digitalen Welt

Von Hans Rütimann

Hintergrundinformation:

Als ich vor drei Jahren an dieser Stelle anläßlich der feierlichen Einweihung des Instituts für Erhaltung von Archiv- und Bibliotheksgut einen Vortrag hielt, hatte ich noch *einen anderen Hut auf* und vertrat die Commission on Preservation and Access (CPA). Diese hat sich inzwischen mit dem Council on Library Resources vereinigt; die zwei Organisationen bilden nun CLIR, den Council on Library and Information Resources. Die Fusionierung hat viel mit einem erweiterten Arbeitskreis zu tun. Die Fusionierung heißt keinesfalls, daß nun die Bestandserhaltungprobleme in Amerika und weltweit zur allseitigen Zufriedenheit gelöst seien und man nun zu neuen Ufern steuern könnte. Auch in der neugebildeten Organisation CLIR hat diese Aufgabe hohe Priorität, aber auch neue Dimensionen: wir haben es inzwischen nicht mehr nur mit endogenem Papierzerfall zu tun, sondern mit der neuen Art von digitalen Informationen, die, wie Jeff Rothenberg festgestellt hat, weit brüchiger sind als brüchiges Papier. CLIR hat die internationalen Aktivitäten ausgeweitet, dabei aber die wesentlichen Ziele des internationalen Programms der CPA beibehalten.

Das internationale Programm

Unsere Arbeit konzentriert sich auf drei Hauptpunkte: erstens über die Notwendigkeit der Bestandserhaltung im Ausland, besonders in den Entwicklungsländern, aufzuklären, zweitens Einrichtungen dabei behilflich zu sein, Methoden und Strategien zu entwickeln, um Probleme des Zugriffs in Bibliotheken und Archiven zu lösen, und drittens, wo angebracht, Unterstützung bei der Ausbildung und anderen Maßnahmen zu leisten, die den Aufbau institutioneller Kapazität ermöglichen, um den Herausforderungen der Bestandserhaltung und des Zugriffs gewachsen zu sein.

Das internationale Programm hat seine Bemühungen bisher auf Ost- und Westeuropa, die ehemalige Sowjetunion, China und Lateinamerika konzentriert. Dank der Unterstützung der Andrew W. Mellon-Stiftung wird die Arbeit in Lateinamerika ausgedehnt, und CLIR wird neue Projekte in Südeuropa (z.B. in Griechenland) und in Südafrika ins Leben rufen. Wir hoffen, das Programm in Asien zu erweitern, und suchen Mittel, um dort neue Projekte zu entwickeln. Zu den Projekten, die in den nächsten Jahren durchgeführt

werden, gehören z.B. die Bereitstellung von Fachwissen über die Evaluierung der Bestandserhaltungsnotwendigkeit in ausländischen Bibliotheken und Archiven, die Entwicklung gemeinsamer Verfilmungs- und Digitalisierungsprojekte für spezifische Sammlungen und die Entwicklung von Strategien, um die Herstellung und Anwendung von alterungsbeständigem Papier zu erweitern.

Ein spezifisches Beispiel:

Wir betreuen ein Projekt *Übersetzung und Verbreitung von Kenntnissen in Brasilien über die Bestandserhaltung.* Zugang zu Information bedeutet oft die Übersetzung von fachspezifischer Literatur über Bestandserhaltung in andere Sprachen. Ein zwischeninstitutionelles Bündnis interessierter Organisationen in Brasilien überwacht das Projekt, zu dem die Übersetzung von 52 Werken über Bestandserhaltung, von der Umweltkontrolle bis zur digitalen Konversion, ins Portugiesische gehörte. Die Übersetzungen bildeten die Grundlage für Arbeitsseminare im ganzen Lande. Dabei sammelten die Projektleiter wertvolle Informationen über den Zustand der Sammlungen in mehr als 1 400 Bibliotheken und Archiven. Vor kurzem erhielt das Projekt nicht nur umfangreiche Unterstützung zur Weiterführung und Erweiterung, sondern wurde auch vom brasilianischen Kultusministerium mit der höchsten Auszeichnung auf dem Gebiet der Erhaltung des Kulturerbes geehrt. Das Projekt wird weiter ausgedehnt. Regionale Koordinatoren in allen brasilianischen Bundesstaaten leiten Arbeitsgruppen und Seminare auf der Grundlage der portugiesischen Übersetzungen und berichten regelmäßig über ihre Erfahrungen an den Hauptkoordinator. Die Übersetzungen sind auch zum Sprungbrett für neue Bemühungen in den portugiesischsprachigen Ländern Afrikas geworden. Vor kurzem boten wir während einer Konferenz in Aveiro, Portugal, den Bibiliothekaren und Archivaren aus lusophonen Ländern die in Brasilien erarbeiteten Unterrichtsmaterialien an. Da in den meisten Ländern der Zugriff auf Fachliteratur über grundlegende Bestandserhaltungsbelange für Originaldruck, -ton und -bild von größter Priorität ist, wurde das Angebot mit Begeisterung angenommen. Die Übersetzungen werden die Grundlage für Seminare und Arbeitsgruppen in den Ländern bilden, in denen das Bewußtsein über die Notwendigkeit der Bestandserhaltung und damit verbundene Belange noch in den Kinderschuhen steckt. In Kap Verde und Mosambik ist eine solche Ausbildung zur Zeit im Gange.

Tradition und neue Technologien

Die Bedeutung der Unterstützung traditioneller Präventivmaßnahmen zur Erhaltung von Druck-, Bild- und Tonbeständen kann nicht genug hervorgehoben werden. Besonders in den Entwicklungsländern besteht die Gefahr, daß man der Digitalisierung als Bestandserhaltungsmaßnahme des Kulturer-

bes eines Landes zu viel Bedeutung beimißt. Bibliothekare und Archivare in den Entwicklungsländern und anderswo sind sich oft nicht darüber im klaren, daß digitale Speicherung für langfristige Archivierung von Information sorgfältige Planung erfordert, daß viele organisatorische und technische Probleme noch nicht gelöst sind und daß die Digitalisierung kein Mittel zur Bestandserhaltung ist, solange kein langfristiger Plan das Überleben digital gespeicherter Information sichert.

Besonders besorgniserregend ist, daß viele Einrichtungen in der allgemeinen Begeisterung über alles Digitale traditionelle und grundlegende Erhaltungsmaßnahmen einschließlich Mikroverfilmung beiseite geschoben haben. Bibliothekare und Archivare in Entwicklungsländern stehen unter erheblichem Druck, *etwas zu digitalisieren*, wobei sie eher eine sogenannte *cottage industry* – isolierte Einzelprojekte – als eine kohärente Strategie innerhalb und unter den Institutionen fördern.

Wir sind dafür verantwortlich, mit unseren ausländischen Kollegen nicht nur unsere Erfolge, sondern auch unsere Probleme und unbeantworteten Fragen auszutauschen, wie Abby Smith, Programmdirektorin für CLIR, in ihrem Beitrag *Warum digitalisieren?*[1] hervorhebt:

Obwohl Digitalisierung manchmal einfach als 'Erhaltung' bezeichnet wird, ist klar, daß der stärkste Punkt digitaler Ressourcen bisher die Verbesserung des Zugriffs auf Information ist, während ihr schwächster Punkt die traditionelle Bibliotheks- und Archiv-Bestandserhaltungsarbeit ist. Da Digitalisierung wie Mikroverfilmung eine Art der Umformatierung ist, wird sie ironischerweise oft mit Mikroverfilmung als Bestandserhaltung verwechselt und, obwohl bis jetzt noch teurer, als überlegenere Art der Umformatierung zur Bestandserhaltung angesehen. Digitalisierung ist jedoch keine Bestandserhaltung. Digitalisierung hat viele Vorteile, aber Dauerhaftigkeit und Authentizität gehören zu diesem Zeitpunkt der technologischen Entwicklung nicht dazu.

Andere Stimmen der Erfahrung müssen in vielen Ländern Gehör finden und unrealistischen Erwartungen an digitale Lösungen entgegengestellt werden, wie z.B. die von Don Waters, dem Direktor unserer Digital Library Federation, der in der neuesten Ausgabe von *CLIR Issues* schreibt:

In Anbetracht der guten Aussichten der sich entwickelnden digitalen Umgebung mit ihren vielen einfach anzuwendenden Funktionen leistete die Bibliothekswelt Anfang und Mitte der neunziger Jahre umfangreiche Forschungsarbeit zu der Frage, ob Digitalisierung als Ersatz für Mikroverfilmung als eine Umformatierungstechnik zur Bestandserhaltung dienen könne.

Nach Durchführung zahlreicher Experimente sammelten sich nur wenige Beweise dafür an, daß die Umstellung auf digitales Format ein angemessener Ersatz für Mikroverfilmung zur Bestandserhaltung ist. Die sich schnell verändernden Technologien haben es den Forschern schwer gemacht, die digitalisier-

---

[1] Abby *Smith:* Why Digitize? Hg. von Council on Library and Information Resources. Washington D.C. 1999.

ten Versionen am Leben zu erhalten, damit die Leser die angeblichen Vorteile des verbesserten Zugriffs nutzen können. Noch viel weniger können sie sich auf solche Versionen als dauerhaftes Bestandserhaltungsmittel verlassen, das den Inhalt von brüchigen Originalen erhält. Deshalb verlagerte sich die Forschung auf einen sogenannten hybriden Ansatz. Dieser Ansatz setzt die ständige Weiterentwicklung von Umweltkontrollen, Entsäuerungen und Mikroverfilmung als Technik für die langfristige Bestandserhaltung von brüchigen Originalen voraus und konzentriert sich auf die Entwicklung verläßlicher und kosteneffektiver Wege der Verwendung von Digitalisierung und anderen digitalen Mitteln als einer Möglichkeit zur Verbesserung des Zugriffs auf diese Materialien.

Don Waters Aussage bedeutet auch, daß die Mittelzuweisung für Bestandserhaltung und Zugriff in Industrieländern wie auch in Entwicklungsländern, die zur Zeit auf digitale Entwicklung konzentriert ist, im Hinblick auf eine ausgewogenere Verteilung, und zwar eine solche, die die Pflege des Originals und die altbewährte Mikroverfilmung zur Bestandserhaltung nicht vernachlässigt, neu überdacht werden sollte. Wenn wir in beiden Richtungen voranschreiten, statt die Bestandserhaltung vom Zugriff zu trennen, bleibt das zugrundeliegende Konzept von *Bestandserhaltung und Zugriff* unversehrt.

Wenn unsere ausländischen Kollegen unter Druck stehen zu digitalisieren, sind sie auch für den folgenden Rat dankbar (Waters in *CLIR Issues*): *Die Fragen, die bei der Auswahl von Material zur Digitalisierung gestellt werden müssen, befassen sich daher nicht hauptsächlich mit der Erhaltung des Originals, sondern mit dem Zugriff: Wie zugänglich ist das Original? Wie hoch ist die Nachfrage nach dem Original? Wie wird Digitalisierung Zugriff und Benutzung beeinflussen? Leider haben die Bemühungen, Digitalisierung zu rechtfertigen, bis heute selten die nüchterne Analyse der Benutzernachfrage eingeschlossen, die notwendig ist, um solche Fragen zu beantworten.*

Die Sonderarbeitsgruppe zur Archivierung digitaler Information (Task Force on Archiving of Digital Information)

Ein wichtiger Beitrag zur Debatte über langfristige Archivierung digitaler Information kam von der Sonderarbeitsgruppe zur Erhaltung digitaler Information, die gemeinsam von der Commission on Preservation and Access und der Research Libraries Group (RLG) gegründet wurde. Wir haben uns sehr bemüht, diesen Bericht an Bibliotheken und Archive weltweit zu verteilen, da er viele Fragen im digitalen Bereich aufwirft. Die Arbeitsgruppe wurde beauftragt, über Möglichkeiten zu berichten, mit denen die Gesellschaft in Bezug auf Aufzeichnungen des Kulturerbes, die zur Zeit in digitalem Format erstellt werden, vorgehen sollte, nämlich nicht nur durch Konversion vom Druck zum digitalen Format, sondern, vielleicht noch wichtiger, durch Aufzeichnungen, die *digital geboren* sind. Man bedenke, daß bis zum Jahre 2000 unge-

fähr 75% aller Transaktionen der U.S.-Bundesregierung elektronisch durchgeführt werden. Wenn Präsident Clinton sein Amt verläßt, wird seine Administration schätzungsweise mehr als 8 Millionen *electronic records* an das Nationalarchiv übergeben; der Großteil dieser Information ist digitalen Ursprungs, und es gibt oft keine gedruckte Aufzeichnung, auf die man zurückgreifen könnte.

Im 1996er Bericht der Sonderarbeitsgruppe steht, daß *das Problem der Erhaltung digitaler Information für die Zukunft nicht nur, und nicht einmal in erster Linie, ein Problem der Feineinstellung eines engen Satzes von technischen Variablen ist*. Es ist eher *ein Problem, wie wir uns im Laufe der Zeit selbst und als Gesellschaft im ganzen einstellen, um in der digitalen Landschaft effektiv zu fungieren. Es ist ein Problem des fast vollständigen Neuaufbaus der unterschiedlichen Unterstützungssysteme oder der tiefgehenden Infrastruktur, der uns in die Lage versetzen wird, unsere Ängste zu überwinden und unsere Aufzeichnungen des kulturellen Erbes natürlich und mit Zuversicht für die Zukunft zu bewahren*.

Ich nehme an, daß die meisten von Ihnen mit dem Bericht der Sonderarbeitsgruppe vertraut sind und werde daher nicht auf Einzelheiten eingehen (er steht auch als Volltext auf der Internetseite von RLG, *www.rlg.org*, zur Verfügung). Der Titel ist *Preserving Digital Information; Report of the Task Force on Archiving of Digital Information.*

Der Abschlußbericht konzentriert sich auf drei Hauptfragen: Was bringt digitale Bestandserhaltung mit sich? Welche Struktur geben wir uns, um sie durchzuführen? Welches sind die nächsten konkreten Schritte? Die Hauptschlußfolgerung des Berichts kann nicht genug betont werden: unsere größten Herausforderungen im digitalen Zeitalter sind nicht technischer, sondern organisatorischer Art. In dem Bericht wird ebenfalls festgestellt, daß uns zur Zeit die Infrastruktur der Verfahrensweisen, Normen und Organisationen fehlt, die notwendig ist, um die Erhaltung digitaler Information zu unterstützen. Inzwischen wird in vielen Ländern einiges getan, um diese Probleme zu lösen und die Notwendigkeit, Information über die besten Verfahrensweisen im breiten Spektrum der verschiedenen Organisationen und Länder auszutauschen, ist überwältigend.

The Digital Library Federation (DLF) – Vereinigung der digitalen Bibliotheken

Der Bericht dieser Sonderarbeitsgruppe bildete eine der Grundlagen für das Programm der Digital Library Federation. Da meine Kollegen mehr Information über DLF präsentieren werden, möchte ich nur einige der Höhepunkte ihrer Arbeit zur langfristigen Archivierung von Information in digitalem Format (sowohl konvertiert wie auch *digital geboren*) aufzeigen.

Die Hauptaufgabe der DLF ist die Schaffung der notwendigen Bedingungen zur Erstellung, Aufrechterhaltung, Erweiterung und Erhaltung einer dezentralen Sammlung digitalen Materials, die Wissenschaftlern und einer breiten Öffentlichkeit zugänglich ist. Die Mitglieder der Vereinigung haben sich zu einer gemeinsamen Investition für die Entwicklung der Infrastruktur, die für Bibliotheken digitaler Werke notwendig ist, verpflichtet. Die Infrastruktur soll digitale Bibliotheken in die Lage versetzen, die Werke, die sie für ihre Leser verwalten, zusammenzubringen oder zu *föderalisieren*.

Die DLF hat folgende Prioritäten für ihr Programm aufgestellt:

- Die DLF konzentriert sich auf Bibliotheken mit *digital geborenen* Materialien. Es ist äußerst wichtig, daß die Bibliothekswelt außer in besonderen und gut begründeten Fällen von der Konversion zur Organisation, zum Zugriff und zur Erhaltung von *digital geborenem* Material übergeht. Eine hohe Priorität hat die Entwicklung der archivarischen Mechanismen, die die Integrität und Benutzbarkeit dieser digitalen Werke über einen langen Zeitraum erhalten. Die Vereinigung lenkt deshalb die Aufmerksamkeit der Bibliotheken in zunehmendem Masse auf die zahlreichen und schwierigen Themen im Zusammenhang mit Werken, die in digitalem Format *geboren* und nicht zu digitalem Format konvertiert worden sind.
- Die DLF wird dazu beitragen, digitale Materialien in die Struktur des akademischen Lebens zu integrieren. Ein kritischer Punkt für eine solche Integration ist die Definition der Umstände, unter denen die Konversion auf digitales Format gerechtfertigt ist. Konversionsprojekte, die die Erweiterung der Universitätsausbildung fördern und versprechen, die Qualität von Forschung und Lehre zu verbessern und deren Kosten zu senken, verdienen besondere Aufmerksamkeit.
- Die DLF wird dazu beitragen, die Entwicklung einer Kerninfrakstruktur für digitale Bibliotheken voranzutreiben. Die größte Aufmerksamkeit muß zu diesem Zeitpunkt auf die Netzwerk- und Systemanforderungen, die Möglichkeiten der Authentizitätsbestätigung und der Autorisierung, der Auffindung und Abfragung und der Archivierung gelenkt werden.
- Die DLF wird dazu beitragen, die organisatorische Unterstützung, die notwendig ist, um digitale Bibliotheken effektiv zu verwalten, zu definieren und zu entwickeln. Zu den organisatorischen Themen, die von Anfang an berücksichtigt werden müssen, gehören die Identifizierung institutioneller Werte und Strategien zur Handhabung geistigen Eigentums in digitalem Format und die Schaffung der Bedingungen zur Entwicklung von Fachkenntnissen, die für die Verwaltung digitaler Bibliotheken notwendig sind.

Die DLF ist bestrebt, die Beständigkeit digitaler Information, sei sie nun digital *geboren* oder durch Konversion geschaffen, zu sichern. Viel ist über die Notwendigkeit geschrieben worden, digitale Information zu dokumentieren, damit sie die Rolle eines *Dokuments* im archivarischen Sinne übernimmt, wie

auch über die Notwendigkeit, digitale Information auf neue Medien umzustellen, um einen Verlust durch Medienverfall und Veralterung zu verhindern. Jedoch hat man sich mit der grundlegenden Frage, wie sichergestellt werden kann, daß digitale Dokumente in Zukunft lesbar und verständlich bleiben, nur relativ wenig befaßt.

Zu diesem Zeitpunkt unterscheidet man praktischerweise folgende Strategien:

1. Kopieren (keine Informationsveränderung, aber eine Veränderung des Mediums)
2. Systemumstellung (Anpassung an neue Hardware und Software)
3. Emulation (neue Plattformen, die vorhergehende nachahmen)
4. Archäologie (Materialien werden nicht bewußt gespeichert und später als wichtig befunden).

CLIR und die DLF beauftragten Jeff Rothenberg, Autor von *Ensuring the Longevity of Digital Documents* (Sicherstellung der Langlebigkeit digitaler Dokumente – (Scientific American, Januar 1995), die zur Zeit in Betracht gezogenen Ansätze zur Sicherstellung der zukünftigen Zugänglichkeit und Lesbarkeit (d.h. Langlebigkeit) von digitalem Material zu untersuchen. Seine vorläufigen Ergebnisse stellen das Konzept der Systemumstellung oder *Migration* in Frage. Da es sich bei der Migration um die Übersetzung aller verschiedenen Typen und Formate (Text, Bilder, Ton, Video, Animation usw.) handelt, ist der Prozeß äußerst arbeitsintensiv. Wenn ein Dokument für die langfristige Speicherung auf der Grundlage einer gewissen Migrationsstrategie akzeptiert wird, ist es nach Rothenberg unmöglich, auch nur grob einzuschätzen, was in der Zukunft getan werden muß, wann dies getan werden muß, wieviel es kosten wird und wie lange es dauern wird, bis das Dokument durch unangemessene Übersetzung verfälscht ist.

Rothenberg legte CLIR einen Entwurf mit dem Titel *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation* (Wie man technologischen Treibsand vermeidet: Entdeckung einer funktionsfähigen technischen Basis für die digitale Bestandserhaltung) vor. Rothenberg konzentriert sich auf die Emulation als Lösung des Problems der digitalen Bestandserhaltung, eine Lösung, die vorhersehbar und nachprüfbar ist und die die wiederholte Übersetzung von Dokumenten, die *unweigerlich zu deren Verfälschung und Verlust führen muß*, ausschaltet. Andere behaupten, daß Migration wenigstens kurzfristig die einzige funktionsfähige Lösung ist. Diese Diskussion wird ohne Zweifel fortgesetzt.

Schlußfolgerung

Vielleicht sollten wir eine fünfte mögliche Strategie für die Bestandserhaltung von Information in digitaler Form hinzufügen, nämlich Gebete. Das hat der Daiho-Tempel des Rinzai Zen Buddhismus in Kyoto getan, als er ein Gebet für verlorene elektronische Information auf seiner Internetseite veröffentlichte *(thezen.or.jp/jomoh/kuyo.html).*

Ernsthaft: Wir müssen uns weiterhin den vielen, oft unrealistischen Erwartungen an *digitale Lösungen* in Entwicklungsländern und sonstwo stellen. Es gibt wirklich viele Lösungen, die wir aus einer Vielzahl von Ansammlungen an Mitteln zur Bestandserhaltung und zum Zugriff zusammenbauen können – eine zur Erhaltung und zur andauernden Zugriffsmöglichkeit auf Materialien im Original, eine andere für Information, die in der digitalen Umgebung *geboren* und/oder in ihr aufgewachsen sind. Wir sind dafür verantwortlich sicherzustellen, daß wir diese Repertoires an Instrumenten vergrößern, wo immer dies möglich ist, und in allen relevanten Umfeldern das Wissen und Know-how entwickeln, um die richtige Lösung aus dem entsprechenden Repertoire unter allen gegebenen Umständen auszuwählen. Oder, wie Dr. Smith in dem bereits erwähnten Artikel *Warum digitalisieren?* folgert:

Es wäre verbohrt, alles auf digitales Format umzustellen, selbst wenn wir dies könnten. Die wirkliche Herausforderung ist, diese analogen Materialien besser zugänglich zu machen, indem wir das mächtige Werkzeug der digitalen Technologie nicht nur durch Konversion, sondern auch durch digitale Auffindungshilfen und verbundene Datenbanken von Suchhilfen benutzen. Digitale Technologie kann sich in der Tat als wertvolles Instrument zur Verbesserung des Lernens erweisen und den Zugang zu Informationsressourcen für die ausweiten, die sie suchen, wo immer sie auch sein mögen, aber nur dann, wenn wir sie als Zusatz zu einem bereits gut sortierten Werkzeugkasten entwickeln und nicht als Ersatz für all die Bestandserhaltungs-Werkzeuge, die Generationen vor uns genialerweise erfunden und vertrauensvoll an uns weitergegeben haben.

Risiken bei der Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut aus konservatorischer Sicht

Von Gerhard Banik

Einleitung

Die Blaubeurener Empfehlungen für die Restaurierung und Konservierung von Archiv- und Bibliotheksgut (1992) nennen unter dem Punkt 1.2 *Konservatorische Sofortmaßnahmen* die Herstellung von Mikrofilmaufnahmen (Schutzverfilmung) Reproduktionen, also etwa Faximiles, Photokopien, Photographien oder eines sonstigen Ersatzmaterials, wie z.B. die Digitalisierung, als die am wenigsten in die Originalsubstanz eingreifende und wirtschaftlichste Schutzmaßnahme für die Benutzung. Derartige Formatkonvertierungen können den nicht auflösbaren Gegensatz zwischen Schutz und Nutzung des in Archiven und Bibliotheken verwahrten Kulturgutes auflösen helfen. In vielen Fällen wird der Benutzer von Archiv- und Bibliotheksgut sicherlich mit einem vorlagengetreuen und qualitätsvollen Ersatzmedium auskommen können. H. Weber (1992) beziffert diese Möglichkeit mit 90% der Fälle, in denen eine gefährdetes Original geschützt im Magazin verbleibt – somit seine Lebenserwartung verlängert wird – und die Information dem Benutzer trotzdem erfolgreich zugänglich gemacht werden kann. Das Ziel, Archiv- und Bibliotheksgut durch Formatkonvertierung zu schützen, läßt sich jedoch nur dann erreichen, wenn diese Maßnahmen ohne Schädigung der Vorlage erfolgen können. Insbesondere im Hinblick auf gebundene Materialien ist dies leider nicht selbstverständlich, denn um eine verzerrungsfreie und scharfe Reproduktion zu erhalten, muß zumeist der Schriftbereich der Buchseiten annähernd in einer Ebene liegen. Bücher, als dreidimensionale Objekte, sind infolge der Bindetechnik nicht immer darauf angelegt, ein Öffnen auf 180° zu ermöglichen. Wird die Öffnung trotzdem erzwungen, so besteht die Gefahr einer mechanischen Schädigung. Wie groß das jeweilige Gefährdungspotential ist, hängt von der Beschaffenheit des jeweiligen Buches und nicht zuletzt von der eingesetzten Reproduktionstechnik ab. Trotzdem muß man feststellen, daß über die Risiken und Grenzen einer Formatkonvertierung von Archiv- und Bibliotheksgut bisher wenig Angaben in der Literatur zu finden sind. Zwar gibt es detaillierte Veröffentlichungen bezüglich der Mindestanforderungen etwa an die Alterungsbeständigkeit und Verarbeitung der verwendeten Ersatzinformationsträger, doch sind zur Frage der Belastungen und der möglichen Schädigung der Objekte durch eine Formatkonvertierung nur wenige

Arbeiten bekannt. Als Beispiele seien die Abhandlungen von Weber (1992) oder von Keimer (1997) genannt. Dabei scheint das mechanische Problem zu eng gebundener Bücher durchaus im Bewußtsein zu stehen, denn nicht umsonst wurden buchschonende Vorrichtungen wie Buchschwingen oder die Prismen-Kamera mit Ungeduld erwartet.

Risiken durch Lichteinfluß

Zu den Belastungen, denen ein Objekt bei der Formatkonvertierung ausgesetzt ist, sind neben mechanischen Einwirkungen besonders der Lichteinfluß unter besonderer Berücksichtigung von Klimafaktoren zu nennen. Die relative Feuchte spielt in diesem Zusammenhang eine wesentliche Rolle. Es soll in diesem Beitrag zunächst der Einfluß der Beleuchtung diskutiert werden. Die Energie elektromagnetischer Strahlung, also des Lichtes, ist in der Lage sogenannte photochemische Prozesse auszulösen, wobei für die Betrachtung ultraviolette und sichtbare Strahlung von Bedeutung sind. Photochemische Prozesse in beleuchtetem Material lassen sich mit Hilfe der Quantentheorie des Lichtes beschreiben. Diese besagt, daß Energie immer nur in Form von Quanten ausgestrahlt und absorbiert wird. Im Fall der optischen Strahlung bezeichnet man diese Quanten als Photonen.

Die Beleuchtung eines Objektes kann man daher als Photonenbombardement beschreiben (Hilbert 1996). Der Term Photon wird wie folgt definiert: Elementarquantum der elektromagnetischen Strahlungsenergie, gegeben durch das Produkt aus dem Plank'schem Wirkungsquantum $\mathbf{h}$ und der Frequenz $\mathbf{F}$ der elektromagnetischen Strahlung ($\mathbf{E} = \mathbf{h} \times \nu$). Die Frequenz ν ist gleich der Lichtgeschwindigkeit $\mathbf{c}$ dividiert durch die Wellenlänge λ ($\nu = \mathbf{c}/\lambda$). Das heißt, die Energie eines Photons ist umgekehrt proportional der Wellenlänge ($\mathbf{E} = \mathbf{h} \times \mathbf{c} / \lambda$). Je kürzer die Wellenlänge, um so größer ist daher die Energie. Die Wahrscheinlichkeit, daß in einer beleuchteten, lichtempfindlichen Substanz molekulare Prozesse ausgelöst werden, wächst mit dem Anteil energiereicher Photonen, der zur Anregung der Moleküle in der bestrahlten Substanz führt. Dadurch werden Folgereaktionen ausgelöst. Die aktive Absorption ist Voraussetzung für photochemische Prozesse.

Generell unterscheidet man bei den photochemischen Prozessen Aufbauprozesse, etwa die Photosynthese, in diesem Zusammenhang aber wesentlich bedeutender sind Abbauprozesse, die an Materialien durch photochemische Reaktionen eintreten. Diese Abbauprozesse bestehen aus Primär- und Sekundärreaktionen. Zur ersteren gehört die Photolyse, die ohne Beteiligung anderer Substanzen immer dann zu einer Veränderung im molekularen Aufbau eines dem Licht ausgesetzten Materials führt, wenn die Energie der Photonen ausreicht, die in den Molekülen vorhandenen chemischen Bindungen zu zerbrechen. Zur zweiten Kategorie gehören Prozesse wie die Photooxidation

und die Photoreduktion, an denen prinzipiell Fremdmoleküle – hauptsächlich Wasser und Luftsauerstoff – beteiligt sind (zitiert nach Hilbert, 1996: 87–88).

In diesem Zusammenhang soll nur auf die Photooxidation eingegangen werden, die als häufiger Abbaumechanismus bei Farbstoffen, Fasern und anderen organischen Materialien auftritt. So wird etwa Indigo durch Photooxidationsprozesse abgebaut, aber auch Cellulosefasern in Papier, Holz oder Textilien, sind von diesem Abbauphänomen betroffen. Die Bedeutung der Photooxidation beim Abbau organischer Verbindungen läßt sich aus der Tatsache erklären, daß der Luftsauerstoff bei diesem Prozeß als starkes Oxidationsmittel eine wesentliche Rolle spielt. Durch Licht angeregte organische Moleküle kommen in Kontakt mit Sauerstoff und Feuchtigkeit und werden auf diese Weise chemisch verändert.

Ein durch Licht angeregtes Farbstoff Molekül (**F***) wird üblicherweise durch ein Perhydroxylradikal (**·HO_2**) oxidiert, das sich wiederum durch die lichtinduzierte Reaktion zwischen Sauerstoff und Wasser bildet [**1–3**].

$F + h\nu \rightarrow F^*$ [1]
$F^* + H_2O \rightarrow FOH + \cdot H$ [2]
$\cdot H + O_2 \rightarrow \cdot HO_2$ [3]

Die direkte Reaktion des Farbstoffs mit Sauerstoff folgt einem anderen Reaktionsmechanismus führt aber zum gleichen Resultat [**4–6**]:

$F + h\nu \rightarrow F^*$ [4]
$F^* + O_2 \rightarrow F + O_2^*$ [5]
$O_2^* + H_2O \rightarrow \cdot HO + \cdot HO_2$ [6]

Das gebildete Perhydroxylradikal (**·HO_2**) ist ein sehr reaktives, stark oxidierendes Reagenz, daß Oxidationsreaktionen sowohl am Farbstoff als auch am Fasermaterial auslösen kann. Sowohl die Farbstoffmoleküle als auch die Moleküle der Fasern können in Form ihrer Radikale **F*** bzw. **Faser`** mit Sauerstoff entsprechend den Gleichungen [**7** und **8**] reagieren:

$F^* + O_2 \rightarrow FO_2^*$ [7]
$FO_2^* + Faser \rightarrow FO_2H + Faser`$ [8]

Soweit es die Cellulosefasern im Papier oder kollagene Proteine in Hautmaterialien (Pergament und Leder) betrifft, führen diese Reaktionen zur Vernetzung der Moleküle und ihrer Depolymerisation. Das Resultat manifestiert sich in einer Verringerung der mechanischen Stabilität, die sich in einer deutlich verringerten Elastizität äußert und nicht reversibel ist (Brill 1980).

Im Fall von lichtempfindlichen Sammlungsgut sind immer Fremdstoffe an Degradationserscheinungen beteiligt, hauptsächlich die Luftfeuchtigkeit und immer auch Luftsauerstoff. Insbesondere die Luftfeuchtigkeit beschleunigt

viele Abbaureaktionen, so daß die Kontrolle dieses Klimafaktors für die Risikoverminderung bei Lichteinstrahlung auf Sammlungsgut von großer Bedeutung ist.

Die Temperatur als weiterer Klimafaktor spielt eine geringere Rolle, soweit ein Temperaturbereich von 18 bis 25°C, also Raumtemperatur, eingehalten werden kann, weil in diesem Bereich die Veränderung von Reaktionsgeschwindigkeiten als verhältnismäßig klein eingeschätzt werden kann. Zusammenfassend läßt sich über das Risiko von Lichteinstrahlung auf Sammlungsgut Folgendes sagen:

Es besteht ein Zusammenhang zwischen der Lichtabsorption eines Stoffes und dem Ausmaß photochemischer Veränderung insofern, als diejenigen Wellenlängen durch das Material besonders stark absorbiert werden, gegen die es eine besondere Empfindlichkeit zeigt.

Das Ausmaß der durch Lichteinstrahlung hervorgerufenen Veränderungen ist durch die Lichtintensität pro Flächeneinheit und von der Bestrahlungsdauer abhängig.

Organische Materialien sind im allgemeinen wesentlich empfindlicher gegenüber Lichteinstrahlung als anorganische Materialien, d.h., daß Pigmente unempfindlicher reagieren als Farbstoffe, wenn man von photokatalytisch wirksamen anorganischen Verbindungen wie etwa Zinkweiß oder Titanoxid in der Anatas Form absieht (Tabelle 1). Die Empfindlichkeit von Pigmentaufstrichen ergibt sich aber zumeist aus dem Vorhandensein empfindlicher organischer Bindemittel, die den Pigmentaufstrich auf der Oberfläche des Informationsträgers fixieren.

BESTÄNDIGKEITEN GEGENÜBER BESTRAHLUNG MIT SICHTBAREM LICHT	MATERIALIEN MATERIALKOMBINATIONEN
Unbeständig	Farbstoffe mit geringer Lichtechtheit (Blau-Woll-Standard <3) • Füllfedertinten • Kugelschreiberpasten • Faserschreibertinten • Farbige Ink Jet Tinten • Papier mit Zinkweißaufstrichen • Papier mit Titandioxidaufstrichen (Anatas vor 1945)
Empfindlich	**OxyCellulose** • Tintenfraß-Schäden • Farbfraß-Schäden **Proteine** • Pergament • Leder • Seide • Manche Eisengallustinten • Farbstoffe mittlerer Lichtechtheit (Blau-Woll-Standard 4–6)
Relativ beständig	**Reine Cellulose** • **Anorganische Farbmittel** *Ausnahmen:* • *Zinkoxid (Zinkweiß),* • *Titandioxid in Anatasform* • **Graphit** • **Traditionelle schwarze** • **Druckfarben** • **Metallauflagen**

Tabelle 1: Beständigkeiten einiger ausgewählter Materialien bzw. Materialkombinationen gegenüber Bestrahlung mit sichtbarem Licht.

Die Lichtbeständigkeit (Lichtechtheit) von Farbstoffen, Pigmenten bzw. Objekten wird im allgemeinen nach dem Internationalen Blau-Woll-Standard (ISO) in Beständigkeitskategorien von 1 bis 8 beschrieben. ISO 1 beschreibt eine äußerst geringe Lichtechtheit, ISO 8 ein gegen Bestrahlung sehr stabiles Material.

MATERIAL	Blauwollstandard ISO
Farbphotographien	1–2
Gelbe und grüne Naturfarbstoffe	1–2
Indigo auf Baumwolle	3
Zinnober, Alizarin (Pigmente)	4–6
Indigo auf Wolle	7
Ultramarin, Graphit, Erdpigmente	8
Silbergelatine Auskopierpapiere	8

Tabelle 2: Lichtbeständigkeitsdaten nach ISO für einige relevante Farbmaterialien und Objektkategorien zusammengestellt (nach Ashley-Smith 1999, S. 241).

Im Bereich Archiv- und Bibliotheksgut findet sich eine große Anzahl von Materialien, die empfindlich gegen Lichteinstrahlung sind. Dazu gehören u.a. Papiere mit hohem Ligningehalt und hoher Azidität, Beschreibstoffe auf Basis synthetischer Farbstoffe sowie Photographien, darunter insbesondere frühe Salzpapiere und Farbphotographien.

Im allgemeinen manifestiert sich der negative Einfluß von Licht auf die genannten Materialien durch Farbveränderungen, also z.B. durch Verbräunung ligninhaltiger Papiere oder Ausbleichen von Farbstoffen. Die Veränderung ist eine Funktion der Bestrahlungsdosis. Die Bestrahlungsdosis ergibt sich aus dem Produkt der Lichtintensität (**I**) und der Bestrahlungsdauer (**t**). Der Abbau oder die Veränderung von Materialien ist eine Funktion der Bestrahlungsdosis und läßt sich mathematisch ausdrücken: $\mathbf{E} = \mathbf{k} \times \mathbf{I} \times \mathbf{t}$. In dieser Gleichung bedeuten **E** das Ausmaß der Veränderung (des Materialabbaus), **k** ist eine Konstante und das Produkt aus **I** und **t** entspricht der Bestrahlungsdosis.

Je nach Energieinhalt des eingestrahlten Lichts ist die Geschwindigkeit der Schädigung lichtempfindlicher Materialien unterschiedlich. Je höher die Strahlungsenergie, d.h. je niedriger die Wellenlänge, um so schneller der Materialabbau. Abbildung 1 stellt die Rate der Veränderung eines Naturfarbstoffes bei unterschiedlichen Wellenlängen dar.

Die Tatsache, daß bei niedrigen Wellenlängen die Schadwirkung an Farbstoffen erheblich größer ist, setzt zwingend voraus, daß Beleuchtungseinrichtungen für Reproduktionsarbeiten mit UV-Filtern ausgestattet werden, um so die Schadwirkung möglichst zu reduzieren. Es sei aber darauf hingewiesen, daß sichtbares Licht für sich eine Schädigung verursacht (siehe Abbildung 1)

und daher die Ausstattung der Beleuchtungseinrichtungen mit UV-Filtern keinen vollständigen Schutz der Objekte gewährleistet. Insbesondere bei empfindlichen Farbstoffen hat das Ausfiltern der UV-Strahlung nur einen geringen Effekt, bei beständigeren Materialien läßt sich die Lebensdauer allerdings um den Faktor 10 verlängern.

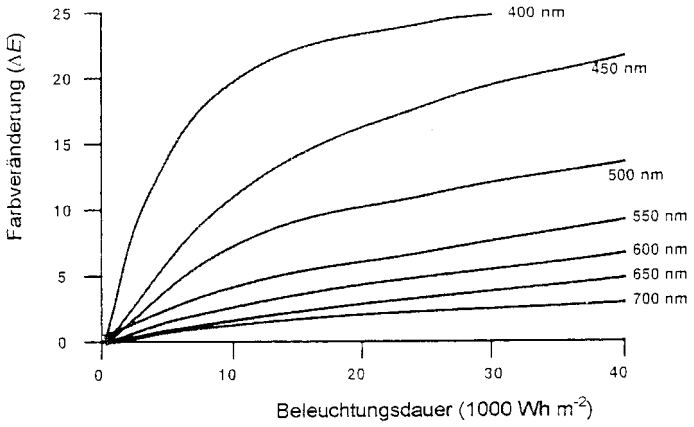


Abb. 1: Bei Reprophotographien mit Blitzlichteinrichtungen beträgt die Beleuchtungsintensität ca. 600 Lux/Sek (Sounders, 1995). Zur Verdeutlichung der Größenordnung kann man festhalten, daß eine Blitzlichtaufnahme bei einem Objekt, das bei einer Beleuchtungsstärke von 50 Lux 8 Stunden täglich ausgestellt ist, der jährlichen Beleuchtungsdosis einen millionstel Teil hinzufügt.

Bei den meisten Techniken der Formatkonvertierung (Verfilmung, Digitalisierung, Herstellung von Photokopien, Reprophotographie) sind die Beleuchtungszeiten kurz. Für einige der genannten Techniken ist der Energieeintrag und die Beleuchtungsdauer in der Tabelle 3 zusammengefaßt.

REPROPHOTOGRAPHIE	MIKROVERFILMUNG	DIGITALISIERUNG
ca. 5 Sek. Belichtungszeit (max.)	ca. 0,5 bis 10 Sek. Belichtungszeit (max.)	ca. 30 Sek. Belichtungszeit (max.)
1000 W Kunstlicht mit UV-Filter	1000 W Kunstlicht mit UV-Filter	1000 W Kunstlicht mit UV-Filter

Tabelle 3: Zusammenstellung von Energieeintrag und Belichtungszeiten für einige ausgewählte Reproduktionstechniken.

Abschätzungen des Risikos durch den Lichteinfluß auf Objekte bei Formatkonvertierungen sind schwierig zu treffen. Es kann festgestellt werden, daß in vielen Fällen die Faktoren, die lichtinduzierte Abbaureaktionen organischer Materialien bestimmen, nicht ausreichend untersucht bzw. verstanden sind. Die meisten Untersuchungen wurden unter kontrollierten, idealisierten Bedingungen durchgeführt und sind für die Bewertung der Materialempfindlichkeit unter realistischen Bedingungen, insbesondere bei kurzzeitiger Lichteinstrahlung, kaum auswertbar.

Indirekte Schlußfolgerungen lassen sich aber aus Untersuchungen von Neevel ziehen, in denen die Gefährdung von Schriftgut und Graphik durch Blitzlichteinrichtungen oder Lichtquellen in konventionellen Photokopierern gemessen wurden. Neevel betrachtete bei seinen Untersuchungen insbesondere den UV-Anteil der beschriebenen Lichtquellen und konnte feststellen, daß im Falle von Blitzlichteinrichtungen im Schnitt 50 Blitze der UV-Dosis einer Kunstlichtbeleuchtung von 50 Lux und einer Stunde Beleuchtungsdauer entsprechen. Bei Lichtquellen in Photokopiereinrichtungen entsprechen im Schnitt 26 Kopien der UV-Dosis, mit der ein Objekt bei einer Kunstlichtbeleuchtung von 50 Lux und einer Stunde ausgesetzt ist.

J. Neevel vermutet, daß die hohe Strahlungsintensität, die von Blitzlichteinrichtungen bzw. Photokopierern abgegeben wird, zu anderen oder wesentlich schnelleren photochemischen Reaktionsmechanismen führen könnte, als dies für geringere Strahlungsintensitäten der Fall ist. Dem widerspricht D. Saunders (1995), der annimmt, daß diese hohen Intensitäten für kurze Bestrahlungszeiten exakt der Wirksamkeit niedriger Bestrahlungsintensitäten für eine längere Bestrahlungsdauer entsprechen. Es sei weiter darauf hingewiesen, daß der UV- Anteil von Photokopiergeräten ausreichend für die Entwicklung von Ozon ist, wobei allerdings die Verweilzeit der Objekte während des Photokopiervorgangs zu kurz für eine Schadwirkung angesehen werden kann.

Unter der Annahme, daß die für die Formatkonvertierung eingesetzten Lichtquellen, also etwa Blitzeinrichtungen oder Scanner, eine vergleichbare spektrale Emission aufweisen, kann daher die Gefährdung der Objekte unter Berücksichtigung der sehr kurzen Belichtungszeiten als gering eingeschätzt werden. Dies auch unter Berücksichtigung der Tatsache, daß etwa Oxycellulose, also z.B. durch Tintenfraß oder Farbfraß stark oxidativ veränderte Cellulose, die sich durch intensive Verbräunung des Papierträgers manifestiert, eine wesentlich höhere Empfindlichkeit gegen Lichteinstrahlung aufweist.

Schädigung durch Manipulation oder mechanische Belastung

Die meisten Schäden, die bei der Formatkonvertierung auftreten können, haben mechanische Ursachen. Sie sind zumeist vergleichbar mit Schädigungen, wie sie auch durch die normale Benutzung von stark abgebauten Papieren oder im Falle von gebundenen Materialien durch Spannungen aufgrund der Bindetechnik beim Aufschlagen hervorgerufen werden. Nicht immer sind diese Schäden auf den ersten Blick erkennbar und können durch partielle Schwächung der Materialien *Sollbruchstellen* schaffen, die in der Folge bei einer Benutzung Schäden hervorrufen. Es lassen sich folgende Belastungs- und Schadensbereiche unterscheiden (siehe auch Tabelle 4):

- Äußerer Einband
- Bindung
- Schriftträgermaterialien, Papier bzw. Pergament
- Beschreib- bzw. Bedruck- oder Malmaterialien

Äußerer Einband

In diesem Fall konzentrieren sich mögliche Schäden auf den Rücken und den Gelenkbereich. Die Materialien reißen ein, brechen oder stauchen. Werden Materialverbünde im Buch gestaucht, so kann dies zur Schichtentrennung führen. Fest gearbeitete Rücken lösen sich, Fehlstellen entstehen, Rückenvergoldungen bröckeln ab, der Ledernarben oder das volle Leder bricht. Besonders gefährdet sind auch die Rücken der meist hohl gearbeiteten Pergamentbände. Das gealterte Material ist häufig nicht mehr elastisch genug, um einer Biegung standzuhalten. Diese Schäden treten besonders häufig bei sogenannten gekälkten Pergamenten auf. Papier- und Gewebeeinbände brechen bevorzugt im Gelenkbereich. Lose Rücken oder auch lose Deckel sind die Folge (Keimer 1997).

Bindung

Zur Bindung gehören die Heftung, die Bünde, Hinterklebungen und die Rundung. Der häufigste Schaden ist die Verformung des Rückens. Gerade Rücken tendieren zur Ausbildung einer Gegenrundung, gerundete Rücken verlieren ihre Rundung, die Rückenform wird unsymmetrisch. Das ist zumeist ein Zeichen für die Lockerung der Bindung insgesamt. Darüber hinaus können Bruchstellen entstehen, die sich bei jeder weiteren Öffnung des Buches markieren und im Extremfall zu einem Auseinanderbrechen des Buchblockes

führen. Bünde aus Leder, aber auch gealterte Heftfäden können brechen oder reißen, so daß der Zusammenhalt von Buchblock und Einband gefährdet ist bzw. sich Blätter oder ganze Lagen aus dem Buchblock lösen. Alle Schäden im Bindungsbereich führen zur Fehlfunktionen, so daß ein schonendes Verfilmen schwieriger wird. Derartige Schäden können zur Zerlegung des ganzen Buchblockes führen. Besonders gefährdet sind Klebebindungen, bei denen sich Schäden besonders schnell und endgültig ausbreiten, denn eine gebrochene Klebebindung ist meistens gleichbedeutend mit dem Auseinanderbrechen des Buchblockes oder seinem Zerfall in Einzelblätter.

Die Schriftträgermaterialien Papier und Pergament

Durch die Manipulation sind, insbesondere durch Alterung, in ihrer mechanischen Stabilität stark beeinträchtigte Papiere gefährdet. Dies gilt besonders für durch Azidität, Farb- oder Tintenfraß mehr oder weniger stark abgebaute Papiere, deren Abbaugrad und mechanische Empfindlichkeit nicht immer visuell erkennbar ist. Folgen der Manipulationen können Risse und Brüche im Schriftbereich, häufig auch Verluste, sein. Pergament reagiert zunächst empfindlich auf Schwankungen der relativen Luftfeuchtigkeit und es liegt für gewöhnlich nicht plan. Bei stark verwellten oder gefalteten Pergamentblättern sollte vor einer Formatkonvertierung mit einem Restaurator abgesprochen werden, ob eine Planlegung des Objektes erforderlich bzw. erreichbar ist. Darüber hinaus sind Beschreibstoffe und Malmaterialien häufig weniger innig mit dem Trägermaterial Pergament verbunden als etwa mit Papier, in dessen Faservlies besonders Tinten oder Farbstoffe besser eindringen können. Daher ist die Gefährdung der Farbschichten oder Beschreibstoffe, bei nicht berührungsfreien Reproduktionstechniken, bei Pergamentobjekten größer als bei Papierobjekten.

Pergamente fanden auch als Umschlagmaterialien in der Bindetechnik Verwendung und weisen, wenn sie in zu trockenem Klima aufbewahrt wurden, nur eine minimale Elastizität auf. Sie neigen daher bei starker mechanischer Belastung zum Brechen.

Beschreibstoffe, Bedruckstoffe, Malmaterialien

Im allgemeinen sind gut fixierte Beschreibstoffe, Druckfarben und auch Farbschichten relativ unempfindlich gegenüber den Manipulationen, die für eine Formatkonvertierung notwendig sind. Problematisch wird jede Manipulation aber dann, wenn durch eine Schwächung des Bindemittelsystems keine ausreichende Bindung von Pigmenten an den Schriftträger mehr gewährleistet ist. Das gilt insbesondere für krakelierte oder pudrige Malschichten, aber auch für

MATERIAL/ MATERIALVERBUND	GEFÄHRDUNG DURCH MANIPULATION
Äußerer Einband	Ablösung von fest gearbeiteten Rücken Brüche des Leders Abbröckeln von Rückenvergoldungen Brüche hohl gearbeiteter Pergamentrücken Hohe Gefährdung *gekalkter* Pergamente durch mangelnde Flexibilität Brüche von Gewebe- oder Papierbänden im Gelenkbereich
Bindung	Rückenverformung Brüche von gealterten Lederbünden Abriß von Heftfäden Bruch von Klebebindungen bis zur völligen Zerlegung des Buchblocks
Papier	Brüche bei: • Tintenfraß-Schäden • Farbfraß-Schäden • Säureschäden • Abgebautem Holzschliffpapier • Harzhaltige Plandokumente vor 1890 • Gealterte Transparentpapiere Risse oder Knicke durch unsachgemäße Handhabung
Pergament	Verwellungen durch zu trockene Raumluft Knicke durch unsachgemäßes Planlegen z.B. durch zu hohen Anpreßdruck
BESCHREIB- UND BEDRUCKSTOFFE, Malschichten	Verluste bindemittelarmer Aufträge bei: Eisengallustinten, Druckfarben, Malschichten
Buch- Archivgut	Verlust von Fragmenten, losen Teilen, eingelegten Zetteln

Tabelle 4: Gefährdungspotential durch mechanische Einflüsse bei Formatkonvertierung (Keimer 1997).

Eisengallus- bzw. Mischtinten oder Druckfarben, die als Folge abgebauter Bindemittel oder von vornherein zu geringer Bindemittelkonzentration, nur eine mangelhafte Bindung an den Schriftträger haben. In diesem Fall kann es bei nicht berührungsfreien Reproduktionstechniken, etwa durch Auflegen einer Glasplatte oder durch das Anlegen an ein Prisma, zum Abheben von Malschichten, Beschreib- oder Bedruckstoffen kommen.

Um vor allem das Risiko von Verlusten von Farbaufträgen, des Transfers von Tinten oder Druckfarben mit mangelnder Bindung zum Schriftträger zu reduzieren, sollten nur berührungsfreie Reproduktionssysteme zur Formatkonvertierung eingesetzt werden. Das bedeutet für die Digitalisierung, daß nur Kamerascanner eingesetzt werden sollten (Tabelle 5).

Ein Digitalisierungssystem für die Bearbeitung von Handschriften, das die erforderlichen Schubsmaßnahmen zur Verhinderung von Bearbeitungsschäden am Sammlungsgut berücksigchtigt, wurde kürzlich an der Universitätsbibliothek Graz entwickelt und vorgestellt. Das System gestattet die berührungsfreie Digitalisierung von gebundenen Handschriften mit einer digitalen Kamera, wobei der Buchöffnungswinkel über eine Buchwippe im Bereich von 120° bis 140° eingestellt werden kann. Die Planlegung des jeweiligen Blattes erfolgt mit einem Saugarm (vgl. technische Daten in Tabelle 6, Mayer, 1999).

NICHT AKZEPTABEL	AKZEPTABEL
Alle Systeme, die Vorlagen berühren: • Handscanner • Einzugsscanner • Trommelscanner • Flachbettscanner*) *) Verursachen Erwärmung bei längerem Betrieb	Berührungsfreie Systeme: • Kamerascanner

Tabelle 5: Ausschlußkriterien für Scannersysteme.

TECHNISCHER VORGANG	DATEN
Beleuchtung	1000 W mit UV-Filter
Beleuchtungsdauer/Blatt	30 Sek. (max.)
Arbeitsdauer/Manuskript Im Durchschnitt	60 min.
Temperaturerhöhung	2°C (max.)
Buchöffnungswinkel	120–140° (in Stufen einstellbar)
Blattfixierung	Saugarm im unbeschriebenen Bereich mit 100 mB Unterdruck

Tabelle 6: Technische Daten der neuentwickelten Digitalisierungseinrichtung für Handschriften an der Universitätsbibliothek (Mayer, 1999).

Zusammenfassung

Unabhängig davon, welche Technik der Formatkonvertierung gewählt wird, ist das entsprechende Objekt einer Reihe von Belastungen ausgesetzt, zu denen neben dem Einfluß der intensiven Lichtbestrahlung, Klimafaktoren wie relative Feuchtigkeit und Temperatur, im wesentlichen mechanische Einflüsse und Probleme der Manipulation gehören. Im Vergleich zu den Risiken der Manipulation können die negativen Einflüsse durch die Beleuchtung als vernachlässigbar gering eingeschätzt werden.

In diesem Zusammenhang ist es wesentlich zu betonen, daß Formatkonvertierungsprogramme nicht durchgeführt werden sollten, ohne daß RestauratorInnen in ein Projektteam eingebunden wird, zu dem BibliothekarInnen, ArchivarInnen oder KuratorInnenen, die technischen BearbeiterInnen, ComputerexpertInnen u.a.m. gehören sollten.

RestauratorInnen sind diejenigen ExpertInnen, die von der materiellen Zusammensetzung des Objektes und seiner eventuellen Gefährdung sicherlich am meisten verstehen, wobei allerdings auch darauf hingewiesen werden soll, daß in vielen Fällen RestauratorInnen nur unzureichend mit den technischen Methoden der Formatkonvertierung vertraut sind.

Literatur

Jonathan *Ashley-Smith:* Risk Assessment for Object Conservation. Oxford 1999. S. 226–245.

Blaubeurener Empfehlungen. Empfehlungen für die Restaurierung und Konservierung von Archiv- und Bibliotheksgut. In: Bestandserhaltung in Archiven und Bibliotheken. Hg. von Hartmut *Weber* (Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg, Serie A, 2). Stuttgart 1992. S. 157–170.

Nicole *Bouché*: Digitalization for Scholarly Use: The Boswell Papers Project at The Beinecke Rare Books and Manuscript Library. Council on Library and Information Resources. Washington D.C. 1999.

Thomas B. *Brill:* Light – Interaction with Art and Antiquities. New York, London 1980. S. 173–195.

Günter S. *Hilbert:* Sammlungsgut in Sicherheit: Beleuchtung und Lichtschutz, Klimatisierung, Sicherungstechnik, #*#Brabdschutz. Berlin [2]1996. S. 72–96.

Barbara *Keimer:* Mikroverfilmung von Büchern als bestandserhaltende Maßnahme (DBI Materialien 154). Berlin 1997.

Manfred *Mayer:* Digitalisierung mittelalterlicher Handschriften an der Universitätsbibliothek Graz, Preprints 9. Internationaler IADA Kongreß Kopenhagen (1999) in Druck.

Johan *Neevel:* Exposure of objects of art and science to light from electronic flash – guns and photocopiers. In: Contributions of the Central Research Laboratory to the field of Conservation and Restoration. In: Central Research Laboratory for Objects od Art and Science. Hg. von H. *Verschoor* und J. *Mosk.* Amsterdam 1994. S. 77–877.

David *Saunders:* Photographic Flash: Thread or Nuisance. In: National Gallery Technical Bulletin 16, National Gallery Publications. London (1995) S. 66–73.

David *Saunders* und Jo *Kirby:* Wavelength-dependent fading of Artist's Pigments. In: Preventive Conservation: Theory and Practice. Preprints to the Ottawa Congress, 12.–16. September 1994, IIC. London (1994) S. 190–194.

Hartmut *Weber:* Verfilmen oder Instandsetzen? Schutz- und Ersatzverfilmung im Dienste der Bestandserhaltung. In: Bestandserhaltung in Archiven und Bibliotheken. Hg. von Hartmut *Weber* (Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg, Serie A, 2). Stuttgart 1992. S. 91–133.

Langzeitspeicherung und Langzeitverfügbarkeit digitaler Konversionsformen

Von Hartmut Weber

Die Erhaltungsproblematik digitaler Informationen

Die Digitalisierung der in den Archiven und Bibliotheken verwahrten Kulturtradition schafft völlig neue Möglichkeiten des ubiquitären Zugangs und ermöglicht neue Qualitäten der Nutzung. Im Vergleich zu den fotografischen und reprografischen Möglichkeiten der Abbildung und Vervielfältigung von Dokumenten, die seit ungefähr 120 Jahren mehr oder weniger zögernd in die Nutzung unikaler Überlieferung eingeführt worden sind, stellt die Digitalisierung einen Quantensprung in den Möglichkeiten der Verbreitung und der Manipulation dar. Das neue und revolutionäre an digitalen Informationen ist der Umstand, daß dabei die Information im Gegensatz zu traditionellem Archiv- oder Bibliotheksgut nicht länger an den Informationsträger gebunden ist. Sie sind einerseits vom Informationsträger gelöst, andererseits aber über Software unlösbar in Funktionalitäten eingebunden und ohne diese nicht lesbar oder nicht verständlich. Die Eigenschaften der digitalen Überlieferungsformen aber, ob sie nun als sekundäre Parallelüberlieferung bei retrodigitalisierten Beständen oder als originäre digitale Überlieferung vorliegt, *digital born* sozusagen, wie dies im amerikanischen Sprachbereich treffend umschrieben wird,[1] lösen einerseits Probleme, schaffen aber auch neue.

Welcher Art diese Probleme sind, bedarf keiner längeren Ausführungen. Schlagworte mögen genügen, um sie zu beschreiben: Abnahme der Intensität der informationsgebenden maschinenlesbaren Signale (Pegelschwund) oder gar deren Störung durch Überlagerung oder deren Verlust. Oxidation, Korrosion oder aus anderen Gründen verursachte chemische oder physikalische Veränderung der Informationsträger wie Magnetbänder oder optoelektronischen Speicher mit entsprechend negativen Auswirkungen auf die auf den Informationsträgern gespeicherten Informationen.[2] Wenn solche chemischen oder physikalischen Veränderungen der Informationsträger irreversibel sind,

---

1 Margaret *Hedstrom* and Sheon *Montgomery:* Digital Preservation Needs and Requirements in RLG Member Institutions. A Study Commissioned by the Research Libraries Group. December 1998. S. 5.

2 Jeff *Rothenberg:* Die Konservierung digitaler Dokumente. In: Spektrum der Wissenschaft (September 1995) S. 66ff.

ist damit die Alterung zutreffend beschrieben. Fast unlösbare Probleme können spontan auftretende Schäden an der Hardware bereiten wie Berührung der Magnetplatten durch die Leseköpfe (*Headcrash*) oder Spannungsspitzen. Die meisten Probleme bereitet schließlich der technische Fortschritt.

Während sich gegen Korrosion und Signalschwund, selbst gegen Totalverluste durch headcrash und möglicherweise auch gegen die Alterung der Informationsträger Vorkehrungen treffen lassen durch wirksame Datensicherung, und redundante Datenspeicherung, steht man dem technischen Fortschritt und seinen Folgen macht- und relativ ratlos gegenüber. Dieser ist gekennzeichnet durch das Veralten der Hard- und Software und der mit zunehmender zeitlicher Entfernung unwahrscheinlichen Eigenschaft der Abwärtskompatibilität von Hardware, Software, Formaten oder Informationsträgern[3], wobei zur Inkompatibilität bereits einer der genannten Faktoren ausreicht.

Auch die Gegenmittel gegen den Verlust von Informationen in digitalen Systemen lassen sich zunächst mit den Schlagworten Migration, Emulation und Konversion beschreiben. Im wesentlichen über diese drei Konzepte zur Langzeitspeicherung und Langzeitverfügbarkeit digitaler Objekte handelt dieser Beitrag. Nicht berücksichtigt wird dabei die irreale Vorstellung, die Langzeitverfügbarkeit digitaler Informationen dadurch zu gewährleisten, daß die erforderliche ursprüngliche Hardware zusammen mit den digitalen Informationen in ihrer Softwareumgebung über lange Zeiträume aufbewahrt und lauffähig gehalten wird. Ferner wird darauf verzichtet, die Möglichkeit zu diskutieren, im Einzelfall mit Hilfe archäologischer Methoden digitale Informationen in obsoleter Umgebung mehr oder weniger aufwendig zu rekonstruieren.

Mit gutem Grund wurde im Realisierungskonzept zur Retrospektiven Digitalisierung von Bibliotheksbeständen im Rahmen der von der Deutschen Forschungsgemeinschaft betriebenen *Verteilten Digitalen Forschungsbibliothek* die Notwendigkeit der *Speicherung zum Zwecke der Langzeitsicherung* unterstrichen.[4] Die Langzeitsicherung ist dabei nicht nur in dem Sinne einer sicheren Speicherung über lange Zeiträume zu verstehen, sondern vor allem im Sinne der Gewährleistung der langfristigen Zugänglichkeit und Nützlichkeit der digitalen Informationen. Diese Notwendigkeit der Zugänglichkeit und des Funktionsnutzens über lange Zeiträume ergibt sich schon allein aus Gründen wirtschaftlichen Denkens und Handelns: die nicht unbeträchtlichen Aufwendungen für Digitalisierungsprojekte sind langfristig abzusichern, da-

---

[3] Charles M. *Dollar:* Die Auswirkungen der Informationstechnologien auf archivische Prinzipien und Methoden. Übers. und hg. von Angelika *Menne-Haritz.* Marburg 1992. S. 97f.

[4] Retrospektive Digitalisierung von Bibliotheksbeständen. Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Farcharbeitsgruppen ‚Inhalt' und ‚Technik' (dbi-materialien 166). Berlin 1998. S. 52f.

mit nicht wiederholt in Projekte investiert werden muß, die dasselbe häufig nachgefragte Archiv- oder Bibliotheksgut zum Gegenstand haben. Unverzichtbar jedoch wird die Forderung der Langzeitverfügbarkeit, wenn nicht der dauerhaften Verfügbarkeit, bei originär digitalen Formen von bleibendem Wert oder in Fällen, in denen digitale Konversionsformen an die Stelle gefährdeter Originale treten sollen oder durch Verlust oder schleichenden Zerfall der Originale früher oder später faktisch treten werden.

Hybridsysteme als präventive Lösung

Digitalisierung vom Mikrofilm

Das bei der Landesarchivdirektion angesiedelte Projekt *Digitale Konversionsformen* ist aus einer intensiven Beschäftigung mit der Frage herausgewachsen, ob denn die Digitalisierung ein adäquates Mittel der Bestandserhaltung sein könne.[5] Dem Ergebnis dieser Voruntersuchung entsprechend, wurde in diesem Projekt von vornherein auf ein Hybridsystem gesetzt, das den Mikrofilm als qualitätvollen aber im übrigen anspruchslosen und vielseitig verwendbaren Langzeitspeicher nutzt, die digitale Konversionsform als attraktive Form für Zugriff, Nutzung und Verteilung. Damit sollen nicht etwa die mit der Langzeitverfügbarkeit digitaler Systeme verbundenen Unsicherheiten und Probleme ignoriert und umgangen werden. Ein Projektteil widmete sich ausdrücklich diesen Fragen. Zum einen lassen sich nämlich auch bei einem hybriden Systemansatz, was die Funktionalitäten der Bereitstellung von digitalisierten Objekten sowie die Kontextinformationen angeht, die Probleme digitaler Erhaltungsprobleme nicht vermeiden, zum anderen sollten Alternativen zum Hybridsystem aufgezeigt und im Wirtschaftlichkeitsvergleich bewertet werden.

Projektverlauf und die Ergebnisse dieses Projekts haben in überzeugender Weise gezeigt, daß die vom Mikrofilm gewonnenen digitalen Konversionsformen sich für alle Zwecke der Bildschirmpräsentation einschließlich angemessener Vergrößerung eignen. Dies gilt für kontrastreiche Schwarzweiß-Aufnahmen auf einem typischen kontraststeigernden AHU-Mikrofilm ebenso wie für Halbtonvorlagen und Photographien auf einem speziellen Halbtonfilm oder für farbige Vorlagen auf einem farbstabilen Farbmikrofilm. Hinsichtlich der Farbübertragung und des dazu erforderlichen Colormanage-

---

[5] Marianne *Dörr* und Hartmut *Weber:* Digitalisierung als Mittel der Bestandserhaltung? Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft. In: Zeitschrift für Bibliothekswesen und Bibliographie Jg. 44 (1997) H. 1 S. 53–76. – Vgl. Hartmut *Weber* and Marianne *Dörr:* Digitization as a Means of Preservation? Washington. October 1997 u. ö.

ments konnten dabei allerdings nur erste Erfahrungen gemacht werden, wobei sich die Probleme der farbtreuen Wiedergabe bei der Digitalisierung vom Film wie unmittelbar vom Original stellen. Im Prinzip macht es auch für weitere und spezielle Schritte der Bildbearbeitung und Verarbeitung keinen Unterschied, ob die Bilddatei unmittelbar vom Original oder über Film gewonnen wurde. Auch in Verbindung mit OCR gibt es keinen signifikanten Unterschied in der in beiden Fällen vorhandenen Fehlerrate. Mit dem Nachweis einer vergleichbar zweckgerechten Qualität ist die wichtigste Voraussetzung für das Hybridsystem aus alterungsbeständigem Mikrofilm und davon gewonnenen digitalen Konversionsformen gegeben.

In einem praktischen Anwendungsfall wurde der Vorteil, vom Film zu digitalisieren, besonders augenfällig. Die im Generallandesarchiv Karlsruhe verwahrte Kurpfälzische Rheinstromkarte aus dem 16. Jahrhundert, ein hochgefährdetes Objekt von ca. 12 m Länge, war vor fast zehn Jahren bereits in Farbe in Teilaufnahmen auf mehreren Makrofiches im Rahmen einer Schutzverfilmungsaktion aufgenommen worden. Das ehrgeizige Ansinnen, für dieses außergewöhnliche Objekt eine Präsentationsform mit den Standardbrowsern am Bildschirm zu entwickeln, konnte ohne erneuten Rückgriff auf das gefährdete Original durch Filmdigitalisierung realisiert werden. Bei der Arbeit mit den Bilddateien ergab sich dann die Notwendigkeit, für die Herausarbeitung von Details dieser von allerhand Personengruppen bevölkerten, mit liebevollen Darstellungen von Schiffen und Flößen und mit detailreich ausgeführten Ortsansichten, vergrößerte Ausschnitte mit angemessener Auflösung zu digitalisieren. Auch dazu reichten die Qualitätsreserven der vorhandenen Makrofiches aus, die dem Dienstleister zum Digitalisieren auf dem Postweg übersandt wurden. Damit konnte an diesem Beispiel nachgewiesen werden, daß zur Schonung des Originals und zur Kostenersparnis vorhandene Filme auch zu einem späteren Zeitpunkt in der jeweils gewünschten Qualität, selbst in Ausschnitten, digitalisiert werden können.

Im Hinblick auf die genannten Vorteile der Hybridlösung stellt sich die Frage nach den Kosten oder den zu vermutenden Mehrkosten für den Langzeitspeicher Film zusätzlich zur digitalen Konversionsform. Eine repräsentative Markterhebung erlaubt in Verbindung mit den eigenen Erfahrungen beim tatsächlichen Einkauf von Digitalisierungsleistungen folgende Größenordnungen zu nennen: setzt man für das Digitalisieren einer Seite im bitonalen Modus vom Mikrofilm einschließlich der Verfilmungskosten 100% an, sind für dieselbe Leistung beim Digitalisieren vom Original 130% zu bezahlen. Bei der Digitalisierung mit 256 Graustufen sind für diese Leistung 330% aufzuwenden, wenn vom Original gearbeitet wird und 195% für die Digitalisierung vom Film. Bei der Farbdigitalisierung sind für das Filmscannen ungefähr 1/3 geringere Kosten anzusetzen als beim Scannen vom Original. Bei schwierigeren Vorlagen spielt dabei auch eine Rolle, daß der panchromatische Mikrofilm mit seinem relativ großen Belichtungsspielraum kontraststeigernd und kontrastausgleichend wirkt, während der Schwellenwert beim Scannen vom Ori-

ginal jeweils individuell und mitunter aufgrund mehrerer Versuche optimal eingestellt werden muß. Die Kombinationslösung aus Mikrofilm und vom Film digitalisierten Konversionsformen ist demnach regelmäßig kostengünstiger als die Digitalisierung vom Original und damit auch wirtschaftlich vorteilhaft.

Die Rekonversion: Ausbelichtung digitaler Konversionsformen

Eine Variante des hybriden Systemansatzes, der im Stuttgarter Projekt verfolgt wurde, nämlich zuerst zu verfilmen und dann vom Film zu digitalisieren, hat insbesondere in den Vereinigten Staaten Anhänger. Diese stellen nicht das Prinzip, den Mikrofilm als Langzeitspeicher einzusetzen in Frage, aber variieren die Reihenfolge von Verfilmen und Digitalisieren und damit den Ausgangspunkt für die analoge und digitale Konversionsform. So wird empfohlen, vom Original zu digitalisieren, um dann von der digitalen Bilddatei ein analoges Bild auf einen Mikrofilm auszubelichten, der dann den Langzeitspeicher darstellt.

Was zunächst wie die beliebte Methode anmutet, bewährte Verfahren zunächst einmal auf den Kopf zu stellen, um mit diesem *Versuchsballon* die Wirkung auf die Fachwelt zu studieren, ist tatsächlich technisch zu realisieren. Mit Hilfe der Technologie des Computer-Outputs on Microfilm (COM) ist es möglich, nicht nur alphanumerische Zeichen (Coded Information CI) sondern auch sogenannte Non Coded Information (NCI) wie digitalisierte Bilder als *bitmaps* auf Film auszubelichten.[6] Eine Kathodenstrahlröhre oder ein Laserstrahl belichtet dabei einen handelsüblichen Mikrofilm, welcher dann konventionell entwickelt und weiterverarbeitet wird. In Deutschland werden auf diese Weise beispielsweise Mikrofiche-Ersatzteilkataloge für die Automobilindustrie mit Bestell- oder Einbauhinweisen und gerasterten Abbildungen hergestellt, die in einer Auflösung um 100 bis 200 dpi in aller Regel bitonal ausbelichtet werden.

Im Rahmen des Stuttgarter Projekts wurden erste Versuche in dieser Richtung zunächst mit Hilfe eines erfahrenen Dienstleisters in Deutschland unternommen. Unter Verwendung einer marktüblichen COM-Anlage wurden Ausbelichtungen mit 24-facher Verkleinerung in einer Bildgröße vorgenommen, die in etwa einem Filmbild auf einem Mikrofiche von 12 x 5 Aufnahmen entspricht. Die Ergebnisse sind von der Wiedergabequalität her zwar nicht mit einem unmittelbar aufgenommenem Mikrofilmbild zu vergleichen. Die Vorlagen sind aber bei Schriftgrößen ab 10 Punkt befriedigend, ab 12 Punkt gut zu lesen. Die Auflösung entspricht 230 dpi, wobei nur bitonal ausbelichtet wer-

---

[6] Vgl. Hybridsysteme: analoge und digitale Speichertechniken im Verbund – AWV Arbeitsgemeinschaft für Wirtschaftliche Verwaltung e.V. Eschborn 1997. S. 10.

den konnte. Trotz dieser eher bescheidenen Wiedergabequalität war mit der Ausbelichtung ein relativ hoher Aufwand verbunden, der nach Mitteilung des Dienstleisters auch bei der Verarbeitung größerer Mengen nicht so weit abgesenkt werden kann, daß diese Dienstleistung als Alternative zur Direktverfilmung unter wirtschaftlichen Gesichtspunkten interessant wäre.

Über einen Untersuchungsbericht der Cornell University Library in Ithaca wurden die Projektbearbeiter auf ein amerikanisches Unternehmen aufmerksam, das COM-Anlagen mit einer speziellen Elektronenstrahl-Technologie entwickelt und entsprechende Dienstleistungen anbietet.[7] Als einziger Anbieter zeigte sich diese Firma (Image Graphics, Shelton CT) in der Lage, mit einem speziellen Electron Beam Recorder EBR in hoher Auflösung von bis zu 600 dpi, sowohl bitonal als auch in Graustufen, digitale Bilddateien im TIF-Format auf handelsüblichen AHU-Mikrofilm auszubelichten.

Die Firma wurde beauftragt, die Bilddateien des *Kleinen Warenkorbs* in den jeweils höchstmöglichen Auflösungsstufen, d.h. 600 dpi bitonal, 400 dpi mit 16 Graustufen (4 Bit) und 400 dpi mit 256 Graustufen (8 Bit) auszubelichten. Bestandteil des *kleinen Warenkorbs* sind drei Testtafeln mit den Auflösungszeichen nach DIN 19051 Nr. 1 (Oktogone) und ISO Nr. 2 (Linienpaare pro mm) sowie Schriftproben mit und ohne Serifen der Schriftgröße 4 Punkt bis 14 Punkt und aufgelegte Graukeile. Die Ausführung des Auftrags zog sich allerdings über Monate hin.

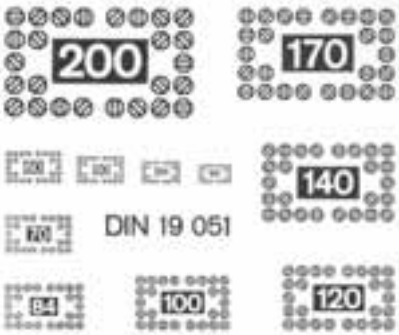


Abb. 1: ISO-Testzeichen Nr. 1.

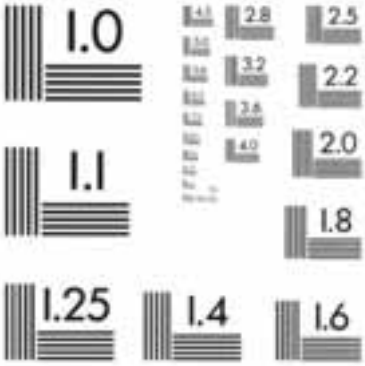


Abb. 2: ISO-Testzeichen Nr. 2.

Der Vergleich zwischen Vorlage, Film, digitaler Konversionsform 1 und digitaler Ausbelichtung sowie davon hergestellter digitaler Konversionsform 2 stellt sich im günstigsten Fall wie folgt dar:

---

7 Anne R. *Kenney:* Digital to Microfilm Conversion: A Demonstration Project 1994–1996. Final Report to the National Endowment for the Humanities PS–20781–94. Ithaca, NY 1997. S. 7ff.

ISO Testzeichen Nr. 2	Auflösungszahl (in Lp/mm)l	Lesbarer Schriftgrad (in Punkt)
Original Testvorlage	18	4
Mikrofilm, AHU oder Halbton	18	4
Duplikat vom Mikrofilm, AHU oder Halbton	14	4
Mikrofilm digitalisiert, 600 dpi, 8 Bit	6,3	4
direkt digitalisiert, 600 dpi, 24 Bit (Konversionsform 1)	8	4
COM Ausbelichtung auf Mikrofilm, 600 dpi, Graustufen 8 Bit	7,1	4
Digitalisierung von der COM-Ausbelichtung, 600 dpi, 8 Bit	6,3	6

Aufgrund der Auswertung von standardisierten Testzeichen lassen sich im möglichst objektivierten Vergleich der Wiedergabequalität analoger und digitaler Formen folgende Feststellungen treffen.[8]

1. Die Lesbarkeit der COM-Ausbelichtung ist subjektiv und objektiv als sehr gut zu bezeichnen, wobei sie jedoch meßbar deutlich (8 Stufen Wiedergabeschärfe weniger nach ISO-Testzeichen) hinter der Wiedergabequalität der Vorlage und der des von der Vorlage unmittelbar auf optischem Wege hergestellten Mikrofilms (ebenfalls 8 Stufen weniger) zurückbleibt.
2. Gegenüber der hochwertigen Konversionsform der Testvorlage, die mit 600 dpi und 256 Graustufen mit einem Buchscanner digitalisiert wurde, tritt nur ein leichter Verlust des Auflösungsvermögens auf (nach ISO-Testzeichen eine Stufe Wiedergabeschärfe weniger).

[8] Das Ausmessen der Testzeichen erfolgte normgerecht mit einem Mikroskop bei 50-facher Vergrößerung. Die Testzeichen auf den digitalen Konversionsformen wurden stets auf dem gleichen IIyama-Bildschirm *Vision Master 500*, Größe 21“, Lochmaske 0,27 bei einer Einstellung von 1280x1024 Bildpunkten (True Color 24 Bit mit der Software ACD-See 32 in der Zoomeinstellung 200% beurteilt.

3. Wird vom digital erzeugten Mikrofilm erneut digitalisiert, ergibt sich gegenüber dem Mikrofilm bei subjektiv und objektiv guter Lesbarkeit aller Bildelemente und aller Schriftzeichen ab 6 Punkt Größe erneut ein leichter meßbarer Schärfeverlust von einer Stufe des ISO-Testzeichens, das Ergebnis ist jedoch vergleichbar mit dem der Digitalisierung vom konventionell hergestellten Mikrofilm.

Im Ergebnis kann festgestellt werden, daß zwar die Wiedergabequalität des ausbelichteten Mikrofilms meßbar deutlich hinter der eines konventionell hergestellten Mikrofilms und auch eines Mikrofilm-Duplikats zweiter Generation zurückbleibt, daß aber die Wiedergabequalität der vom ausbelichteten Film gewonnenen digitalen Konversionsform mit der vergleichbar ist, die von dem auf optischem Wege erstellten Mikrofilm erzeugt wurde. Einschränkungen sind zu erwarten, wenn mit höheren Auflösungen digitalisiert werden soll. In diesem Fall hat der digital ausbelichtete Mikrofilm keine Qualitätsreserven mehr. Wichtig auch die Erkenntnis, daß der digital-analog-digitale Prozeß an den Generationsverlusten teilnimmt, die aus dem analogen Bereich bekannt sind und von Generation zu Generation Kopierverluste aufweist. Dies bedeutet auch, daß Duplikate vom ausbelichteten Mikrofilm schlechter sind, als der Ausgangsfilm und daß Digitalisierungen von Stufe zu Stufe an Qualität verlieren. Dabei ist zu beachten, daß die Wiedergabequalität des digital ausbelichteten Mikrofilms von vornherein 8 Qualitätsstufen unter dem traditionell hergestellten Mikrofilm liegt. Von einer Kompatibilität ‚Film – digitale Form – Film' oder ‚digitale Form – Film – digitale Form' und damit vom Traum eines informationstechnischen Perpetuum Mobile sind wir also nach wie vor noch weit entfernt.

Wenn digitale Konversionsformen von Archiv- oder Bibliotheksgut vorliegen, bei dem die Bildelemente, auf die es ankommt, nicht zu diffizil sind und farbige Langzeitspeicherung und Wiedergabe nicht erforderlich ist, ist das dargestellte Verfahren grundsätzlich anwendbar, um das Problem der Langzeitspeicherung zu lösen. In allen Fällen, in denen Digitalisierungsprojekte neu geplant oder angegangen werden, sollte aber die Frage der Wirtschaftlichkeit darüber entscheiden, ob zuerst verfilmt und dann vom Mikrofilm digitalisiert wird oder ob zuerst digitalisiert und dann auf Mikrofilm ausbelichtet wird. Bisher spricht das Kostengefüge eher für den traditionellen Weg.

Nicht gangbar hat sich der Weg erwiesen, mit einem marktüblichen Diabelichter eine befriedigende Ausbelichtung von farbigen Bilddateien (24 Bit) zu erzielen. Obwohl der für die Tests verwendete Farbfilmbelichter eine Auflösung von 4096 x 2732 Bildpunkten aufwies, waren die Ergebnisse bei Dokumenten und Unterlagen, bei denen es auf eine detailreiche Wiedergabe kleinerer graphischer Elemente ankommt, eher enttäuschend. Die Meßergebnisse bestätigten den subjektiven Eindruck. Auf dem ISO-Testzeichen waren erst die Auflösungsstufe 1,4 und Schriftgrößen ab 14 Punkt lesbar, d.h. die Auflösung lag 21 Stufen unter der des Originals und war 12 Stufen schlechter als der mit dem EBR-Verfahren ausbelichteten Mikrofilme. Einen nicht geringen An-

teil an diesem unbefriedigenden Ergebnis kommt dabei den chromogenen Diapositiv-Filmen zu, die allein in Verbindung mit solchen Geräten benutzt werden können. Diese lösen aufgrund ihrer mehrschichtigen Filmtechnologie von vornherein Schrift und andere graphische Elemente deutlich schlechter auf als Mikrofilme. Im Vergleich bleibt eine Ektachrome-Direktaufnahme des ISO-Testzeichens deutliche 10 Stufen hinter einer Mikrofilmaufnahme zurück. Diese Nachteile des chromogenen Films wirken sich selbstverständlich auch im Diabelichter aus.

Konversionsformen im digitalen Kontext

Die bisher diskutierte Möglichkeit der Langzeitspeicherung digitaler Objekte mit Hilfe der Konversion oder der Rekonversion und der Umgehung des Problems durch einen hybriden Systemansatz kann nur dann vollständig und als alleinige Form zur Anwendung kommen, wenn mit dem Digitalisierungszweck Ziele der Bestandserhaltung zumindest mit verfolgt werden und es damit vorrangig auf den Erhalt der Bildinformation ankommt, oder wenn die digitalen Konversionsformen mit minimalen Kontextinformationen verknüpft werden. In allen anderen Fällen, insbesondere wenn die Qualität der Online-Präsentation digitaler Konversionsformen durch einen Mehrwert an Erschließung und Kontextinformation gegenüber der digitalisierten Vorlage gesteigert wird, werden Überlegungen zur Langzeitverfügbarkeit der Informationen erforderlich, in welche das digitalisierte Objekt eingebettet ist. Um Lösungen für die Erhaltung digitaler Informationen kommt man also um so weniger herum, je mehr informatives und attraktives elektronisches Beiwerk sich um eine digitale Konversionsform rankt.

Dies wird bereits am Modell einer Internet-Bereitstellung deutlich. Das Mindesterfordernis für die Bereitstellung eines digitalen Objekts, beispielsweise einer Photographie oder eines Aktenschriftstücks, zur Nutzung im Internet mit Hilfe eines Browsers ist, daß es identifiziert wird, gezielt aufgerufen werden kann, auf dem Bildschirm an einer gewissen Stelle und in einer gewissen Größe angezeigt wird und daß es schließlich wieder verlassen werden kann. Das Bild liegt dabei in einem komprimierten Format wie JPEG oder PNG vor. Es wird durch Befehle aufgerufen, positioniert, angezeigt, beschrieben und identifiziert, die in der HyperText Markup Language (HTML) codiert sind. Metadaten und Kontextinformationen können so ebenfalls mit dem Objekt verbunden oder in Dateien zur Verfügung stehen, die mit der Web-Seite, auf der das Objekt bereitgestellt wird, mit Hilfe sog. *links* in Verbindung stehen. Der Browser im Endgerät des Nutzers interpretiert diese Auszeichnungs- oder Programmiersprache. Er zeigt nach Aufruf des Objekts dieses in seiner, durch die HTML-Befehle vorgegebenen, Umgebung in der programmierten Darstellung mit den Meta- und Kontextinformationen und mit den vorgegebenen Verbindungen zu anderen Dateien an.

Die dargestellte Architektur einer Bereitstellung digitaler Konversionsformen im Internet wirkt sich auf die Konzepte, die für die Sicherung der Langzeitverfügbarkeit digitaler Informationen zu diskutieren sind, in unterschiedlicher Weise aus. Mit der bereits dargestellten Methode der Konversion können allenfalls die digitalen Objekte konvertiert oder durch Hybridsysteme im Hintergrund abgesichert werden, nicht aber die Metadaten und die Kontextinformationen. Ein Hybridsystem auch bei der Langzeitspeicherung zu unterhalten, kann dennoch zweckmäßig und wirtschaftlich sein, da die Kontextinformationen und Metadaten im Vergleich zur Bilddatei, auch wenn sie komprimiert ist, weitaus weniger Speicherkapazität beanspruchen und, wie anschließend dargestellt wird, nur mit Mitteln der Auffrischung und möglicherweise Formatanpassung fortgeschrieben werden müssen.

Die mit der HyperText Markup Language (HTML) zu einer Bilddatei erzeugten Präsentations- und Steuerungsinformationen liegen im Quellcode vor und werden erst im Browser des Nutzers für die Bildschirmdarstellung und Funktionalität der Präsentation interpretiert. Dies ist das erste im Hinblick auf Langzeitverfügbarkeit bemerkenswerte Faktum. In der zu interpretierenden Programmiersprache lassen sich Funktionalitäten abbilden, eine zweite wichtige Feststellung. Zum dritten ist festzuhalten, daß ein HTML-Dokument ein Datenformat besitzt, das sowohl maschinenlesbar ist als auch für das menschliche Auge lesbar und mit nicht allzu großem Aufwand auch durch Kundige interpretierbar ist. Die 1989 erstmals definierte und heute in der vollständig abwärtskompatiblen Version 4 vorliegende HTML ist eine Anwendung (Dokumententyp-Definition oder DTD) der 1986 international normierten und weit verbreiteten Dokumentenbeschreibungssprache SGML (Standard Generalized Markup Language).[9] In logisch nachvollziehbarer Weise und unter Verwendung englischer *natursprachlicher* Begriffe markiert diese Beschreibungssprache, verkürzt ausgedrückt, alle Elemente eines Dokuments und kann das Dokument und seine Elemente mit beliebigen anderen Dokumenten oder Elementen anderer Dokumente verbinden. Damit kann nicht nur die graphische Darstellung eines Dokuments (oder ganzer Bücher) plattformunabhängig und portierbar gesteuert werden, sondern das Dokument mit beliebigen anderen Informationen verbunden werden, die Bestandteil von entsprechend codierten Dokumenten sind. Dies können auch Metadaten oder Kontextinformationen sein. Entscheidend ist, daß mit Hilfe der SGML-Sprachfamilie ein Dokument mit seinen Funktionalitäten als Einheit dargestellt werden kann, maschinenlesbar und zugleich auch menschenlesbar.

Solange Kontext und Funktionsumfeld eines digitalen Objekts in der im ganzen Internet zwangsläufig verstandenen Sprache HTML beschrieben werden, die bisher konsequent abwärtskompatibel erweitert wurde, ist mit einer

---

[9] Vgl. James *Coleman* and Don *Willis:* SGML as a Framework for Digital Preservation and Access. Washington 1997. S. 9ff.

Langzeitverfügbarkeit im Sinne einer Langzeitverständlichkeit dieser Begleitinformationen zu rechnen. Problematisch sind proprietäre Abweichungen wie das zu Steuerungszwecken von Präsentationen eingesetzte Javascript, das nicht zur SGML-Sprachfamilie gehört oder das Dynamische HTML mit nicht dem SGML-Standard folgenden Erweiterungen zur dynamischen Gestaltung von Web-Seiten, das sich wiederum der Scriptsprachen bedient. Die Kompatibilitätsprobleme von Javascript und Dynamic HTML wirken sich derzeit schon browserabhängig und versionenabhängig aus, was befürchten läßt, daß proprietäre Entwicklungen solcher Art, die sich nicht durchsetzen, früher oder später gnadenlos abgehängt und damit obsolet werden.

Große Chancen wird als künftiger Funktionsausweitung des Standards der Beschreibungssprachen der erweiterten Beschreibungssprache XML Extended Markup Language eingeräumt, die als Teilmenge von SGML näher am ursprünglichen Standard angesiedelt ist, aber bemerkenswerte Funktionalitäten bis zur Abbildung von Datenbankstrukturen aufweist. Die Abwärtskompatibilität zu HTML ist aber auch bei XML-Systemen grundsätzlich gegeben.

Die Langzeitverfügbarkeit digitaler Objekte im Kontext wird demnach wesentlich von der Verwendung standardisierter und weit verbreiteter nichtproprietärer Elemente abhängen. Dies trifft übrigens auch für Präsentationsformate der Bilddateien wie JPEG oder PNG zu. Auch dabei darf nicht auf proprietäre Lösungen gesetzt werden, die im Browser nur in Verbindung mit Plug-ins darstellbar sind. Setzt man auf Langzeitverfügbarkeit (und auf Akzeptanz in der Gegenwart), sollte man sich generell an den gängigsten Browsern im Auslieferungszustand ohne Zubehör und zusätzlich erforderliche spezielle Einstellungen ausrichten. Es bedarf keiner gesonderten Erwähnung, daß man heutzutage ohnehin nur auf Internet Standards und Quasi-Standards setzen darf, auch wenn die digitalen Objekte in lokalen Netzen oder gar vorübergehend auf CD-ROM angeboten werden sollen.

Nach der Analyse, der für den Browser des Endnutzers im Internet erforderlichen Datenstruktur und ihrer Konsequenzen für die Langzeitverfügbarkeit, seien nun zwei typische Möglichkeiten dargestellt, um digitale Objekte und dazugehörende Metadaten und Kontextinformationen dem Endnutzer zur Verfügung zu stellen. Auch dabei ergeben sich unterschiedliche Konsequenzen für die Verfügbarkeit über lange Zeiträume.

Das ebenso einfache wie funktionstüchtige Modell ist das sogenannte Dateisystem (file system). Auf dem Webserver sind die Dateien mit den digitalen Objekten und die Dateien, welche Kontextinformationen oder Metadaten enthalten oder auf die sonst lokal von der Anwendung zugegriffen werden muß, in der Form gespeichert und zugreifbar, wie sie vom Browser des Nutzers abgerufen und interpretiert werden. Das Dateisystem benutzt zur Ablage der Dateien die übliche Baumstruktur, was die Übersichtlichkeit auch bei einer Vielzahl von Dateien gewährleistet. Ein solches Dateisystem kann manuell oder komfortabler mit Hilfe von HTML-Editoren aufgebaut werden. Es kann aber auch automatisch aus einer Datenbank generiert werden. Ein gutes

Beispiel für eine rationelle Umsetzung von Datenbankinhalten in direkt ausführbare HTML-Anwendungen stellt das durch die Deutsche Forschungsgemeinschaft geförderte Projekt *Online Findbuch* der Archivschule Marburg dar.[10] Ein im Rahmen dieses Projekts entwickelter Generator wandelt in erstaunlicher Geschwindigkeit und ohne weiteres Zutun in einer Datenbank gespeicherte Erschließungsinformationen des von der Landesarchivdirektion und von der Archivschule gemeinsam entwickelten Systems MIDOSA95 in ein HTML-Dateisystem um, das den für Online-Anwendungen zu erwartenden Mehrwert enthält, ohne die erforderlichen Fachkonzepte der strukturierten Erschließung von Archivgut aufzugeben. Die Weiterentwicklung der Werkzeuge von MIDOSA Online zur Bereitstellung von digitalem Archivgut wäre ein weiterer wichtiger Schritt.

Das alternative Prinzip zum Dateisystem, das die für die Bereitstellung erforderlichen Dateien sozusagen tafelfertig auf dem Webserver ablegt, ist ein System, bei dem die digitalen Objekte, die Kontextinformationen und die Metadaten in einer Datenbank abgelegt sind und ein Dokumenten-Management-System dafür sorgt, daß auf entsprechende Anforderung über das Netz das jeweils gewünschte Objekt mit den jeweils gewünschten Kontextinformationen *on the fly* in HTML-Dateien, mit eingebetteten digitalen Objekten, konvertiert und bereitgestellt werden. Andere Datenbank-basierte Lösungen verwalten die Kontext- und Suchinformationen in einem Dokumenten-Management-System und die digitalen Objekte auf einem Objekt-Server, wo sie auch nur als Dateisystem abgelegt sein können. Der Webserver ist dabei nur für den Austausch von Anfrage und Antwort in der für jede Seite verständlichen Sprache zuständig. Solche Systeme haben unbestreitbar Vorteile hinsichtlich der Kapazitäten, der Flexibilität und der Interaktivität. Diese Vorteile werden aber dadurch erkauft, daß die Funktionalität von einem sehr komplexen Zusammenwirken von Hard- und Softwarekomponenten abhängt, was eine Langzeitverfügbarkeit solcher Anwendungen in ihrer Funktionalität eher problematisch erscheinen läßt.

Auf die Frage, auf welche Weise digitale Konversionsformen im Kontext ihrer Internet-Präsentation über lange Zeiträume gespeichert und verfügbar gehalten werden können, scheint es zunächst eine globale Antwort zu geben: Wenn sich weltweit eine Institution um die Langzeitspeicherung des weltweiten Internet kümmern würde, müßten nicht an vielen Stellen Überlegungen dazu angestellt und Aufwand getrieben werden. In der Tat gibt es in den Vereinigten Staaten, etwas außerhalb der Stadt San Francisco, das *Internet Archive* als not-for-profit-Organisation des Unternehmens Alexa-Inter-

---

10 Detlev *Heiden* und Mechthild *Black-Veltrup:* Das Marburger Online-Findbuch. Konsequenzen für die Erschließung und Präsentation von Archivgut. In: Der Archivar 52 (1999) Sp. 217ff.

net.[11] Dieses Unternehmen betreibt eine Art Suchmaschine für das Internet und finanziert mit seinen Werbeeinnahmen auch die nicht-kommerzielle Seite, also Internet Archive. Das Internet Archive erstellt im Abstand von ungefähr 60 Tagen eine Schnappschußaufnahme des Internet und hat seit 1996 auf Datenbändern von insgesamt 12 Terrabyte 5 Abzüge des Internet erstellt. Die *Crawler* des Internet Archive erfassen dabei aber nur statisch abgelegte Webseiten, also Dateisysteme. Nicht erfaßt werden dabei PDF-Dateien oder mit Hilfe von Dokumenten-Management-Systemen aus Datenbanken generierte Seiten. Die jährlichen Kosten belaufen sich auf ca. 1 Million US $ und es ist fraglich, wie lange die Firma, die sich um öffentliche Unterstützung etwa durch die Library of Congress bemüht, dieses Geschäft weiter betreibt oder betreiben kann. Neben dem Umstand, daß Internet Archive Bilddateien aus dem Internet allenfalls in komprimierter Form aufzeichnet und daher für eine Speicherung digitaler Master zu Substitutionszwecken ausscheidet, hat der Nutzwert des Internet Archive eine sehr wesentliche Einschränkung. Es findet keinerlei Erschließung statt. Einem potentiellen Nutzer bliebe nichts anderes übrig, als pro Schnappschuß des Internets 2 Terrabyte Daten in Portionen von 35 Gigabyte mit geeigneten Suchmaschinen zu durchkämmen. Die Sorge um die Langzeitverfügbarkeit ihrer digitalen Konversionsformen im Kontext kann das Internet Archive den betroffenen Institutionen offensichtlich nicht abnehmen.

Migration

Langzeitverfügbarkeit digitaler Information durch kontinuierliche Migration zu erreichen, ist der eher traditionelle Ansatz. Durch die Migration soll erreicht werden, daß die Text- oder Bilddateien mit ihren Kontext- und Erschließungsinformationen über sehr lange Zeiträume im Umfeld der jeweils zeitbezogenen Hard- und Softwareumgebungen verfügbar und lesbar bleiben. Dies geschieht einerseits durch Auffrischung der Signale, andererseits durch kontinuierliche Anpassung der Aufzeichnungsformate (*refreshment and reformatting*) und nicht zuletzt durch Anpassung an die geänderte Software, welche die Informationen und die mit ihnen verbundenen Funktionalitäten zugänglich erhält.[12]

---

11 Brewster *Kahle:* Archiving the Internet, erweiterte Fassung eines Artikels in Scientific American 3/1997 unter URL: *http://archive.org/sciam_article.htm.* – Vgl. Ewald *Brahms:* Alexa Internet, San Francisco. In: Elektronische Bibliotheken in den USA. Berlin 1999. S. 160f.

12 Ein Beispiel einer erfolgfreichen aber aufwendigen Migration, welche die Umsetzung von Informationen aus obsolet gewordenen Lochkartensystemen in ein zukunftsoffenes Datenformat umfaßte, wird geschildert bei Ann *Green*, JoAnn *Dionne* and Martin *Dennis:* Preserving the Whole. A Two-Track Approach to Rescuing Social Science Data and Metadata. Washington 1999. S. 6ff.

Ein planmäßiges Migrationskonzept könnte so aussehen, daß zunächst eine Sicherung auf zwei gegeneinander und mit der Datenquelle geprüften zuverlässigen und nicht löschbaren Informationsträger vorgenommen wird. Rechtzeitig vor einem Systemwechsel müßte dann eine Parallelinstallation der beiden Systemgenerationen erfolgen, bis der Schritt der Datensicherung im Umfeld der neuen Systemgeneration erreicht ist. Damit wären dann die reinen Informationen, die Texte oder die Bilddateien gesichert, nicht aber in jedem Fall die Funktionalität der ursprünglichen Anwendung. Bei der Internet-Präsentation eines Buches wäre nicht einmal in jedem Fall sicherzustellen, daß der Zusammenhalt der Seiten und die Verknüpfung mit Metadaten und Kontextinformationen bei der Migration erhalten bliebe.

Die Bereitstellung einer digitalen Konversionsform, beispielsweise einer Buchseite im Internet, macht augenfällig, welches komplexe System dahinter stecken kann. Werden die Bilder, Metadaten und die Kontextinformation in Datenbanken verwaltet und bei Bedarf über ein Datenmanagementsystem einem Webserver als *on the fly* generierte Webseite bereitgestellt, sind bei der Migration nicht nur die Bilddaten, sondern auch die Metadaten und Kontextinformationen, die Software zur Verwaltung und Bereitstellung der Informationen sowie die Datenbank und die Datenbanksoftware zu migrieren, um eine 1 : 1 Funktionalität für die Zukunft zu erhalten. In der Praxis müßte man auf der neuen Plattform ein System mindestens mit der Funktionalität des bisherigen Systems finden und die ganze Anwendung auf dieses anpassen.

Bei der Migration erscheinen das *Refreshment* und das *Reformatting* eher unproblematisch. Im Vergleich zur Migration von Funktionalitäten trifft dies zweifellos zu. Aber auch die Signalauffrischung und Formatanpassung werden in der Regel unterschätzt. Diese unverzichtbaren Arbeitsschritte erfordern einen Aufwand, der mit den relativ geringen Kosten von neuen Informationsträgern nicht abgetan wäre. Dazu kommen die deutlich höheren Kosten für die Inanspruchnahme von Hardware, denn alle diese Prozesse laufen über die Zentraleinheit. Und schließlich fallen die Personalkosten für die Kopier- und Verarbeitungsvorgänge sowie für die Qualitätskontrolle am meisten ins Gewicht. Schreibvorgänge auf Magnetkassetten oder optoelektronische Medien bleiben potentiell fehlerbehaftet. So müssen, zumal bei Migrationsvorgängen, alle neu beschriebenen Informationsträger mit dem Kopiermaster gegengeprüft werden. So entstehen allein bei der Auffrischung der Signale Kosten, die beispielsweise bei einer beschreibbaren Compact-Disc (CD-R) in der Größenordnung von 40 bis 60 DM liegen können.

Sehr viel aufwendiger in jedem Fall, als die Auffrischung von Daten auf Informationsträgern oder die Formatanpassung an neue Systemumgebungen, um die Lesbarkeit von Daten zu gewährleisten, ist die Migration von Funktionalität. Die Übertragung von Funktionalitäten ist im Ergebnis auch mit sehr viel Unsicherheiten und Risiken belastet. Neuerlich hat der in dieser Richtung als Vordenker anerkannte Jeff Rothenberg von der RAND Corpo-

ration erhebliche und begründete Zweifel angemeldet, daß die Migration in der Praxis überhaupt verläßlich funktionieren kann.[13] Nach seiner Meinung ist das Migrationskonzept für die Lösung des Problems insbesondere aufgrund der Unmöglichkeit, die technische Entwicklung zu prognostizieren, ebenso ungeeignet wie die Einrichtung von Hardware-Museen. Falsch ist in jedem Fall ein blindes Vertrauen, daß die künftige Technik schon alles richten werde. Dieses blinde Vertrauen ist auch deshalb für den Fortbestand der Informationen gefährlich, da es Probleme und Risiken, aber auch die Vorsorge aus der Sicht der Gegenwart, ganz in die Zukunft verlagert. Zweifellos ist anzunehmen, daß kommerzielle Anwendungen, die Gewinn abwerfen, solange sie das tun, auch über längere Zeiträume in ihrer Funktionalität ständig fortgeschrieben werden. Ob dies aber auch für Digital Libraries und für entsprechende Angebote der Archive gelten kann, die von der öffentlichen Hand subventioniert werden, ist eher zweifelhaft.

Emulation

Jeff Rothenberg setzt sich alternativ zu dem von ihm als zu risikoreich und unsicher abgelehnten Migrationskonzept für das Emulationskonzept ein.[14] Im Gegensatz zur Migration von Daten und Funktionalitäten der Anwendung, die in die Zukunft gerichtet ist und bei jedem Systemwechsel oder beim Wechsel wichtiger Systemkomponenten erfolgen muß, ist das Ziel der Emulation, die Funktionalität eines Systems in der Zeit, in der es entstanden ist, nachzubilden. Das Emulationskonzept ist von der Zukunft in die Vergangenheit gerichtet. Das Bestechende am strategischen Ansatz der Emulation ist, wie erwähnt, ihre Zielrichtung in die Vergangenheit, über die man in der Regel besser Bescheid weiß, als über die Zukunft. Künftigen Systemen soll es möglich gemacht werden, die Systemvergangenheit zu emulieren. Dadurch soll eine nicht mehr vorhandene Hardware- und Betriebssystemumgebung nachgeahmt werden, um die digitale Information in ihrer ursprünglichen Softwareumgebung und damit auch in der ursprünglichen Funktionalität zugänglich zu machen und zu halten.

Dazu ist es in der Praxis erforderlich, neben den digitalen Konversionsformen die digitalen Kontextinformationen, die Anwendungssoftware und die

---

13 Jeff *Rothenberg:* Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. A Report to the Council on Library & Information Resources (CLIR). Washington 1999. P. 12ff.

14 Ebd., p. 15ff. – Vgl. Jeff *Rothenberg:* Metadata to Support Data Quality and Longevity. 1996. URL: *http://www.computer.org/conferen/meta96/rothenberg_paper/ieee.data-quality.htm*

Betriebssystemsoftware, auf denen diese Software lief, zusammen mit einer ausreichenden Dokumentation der Hardware und des Betriebssystems, einer Beschreibung der Anwendungssoftware und einer Beschreibung, wie alle diese Komponenten zusammenspielen, dauerhaft mit Hilfe der periodischen Auffrischung der digitalen Informationen und Speicherung auf den jeweils gängigen Informationsträgern zu stabilisieren und regelmäßig fortzuschreiben. Für einen Teil der Software enthält also auch das Emulationskonzept die regelmäßige Auffrischung der Daten (refreshment). Mit Hilfe der Dokumentation und einem gewissen Aufwand sollte man mit dieser Methode Anwendungen der Vergangenheit mit eingebetteten digitalen Konversionsformen auf Systemen nächster, übernächster und künftiger Generationen überhaupt zum Laufen bringen, so wenigstens die Vorstellung. Gegenüber der Migration hat die Emulation den Vorzug, daß die langfristige Zugänglichkeit vom jeweiligen Bedarf abhängig gewährleistet werden könnte und nicht digitale Informationen unabhängig vom Bedarf von Generation zu Generation transformiert werden müßten. Der Hauptvorteil, da zugleich ein Kostenvorteil, liegt darin, daß Emulatoren für jedes Betriebssystem nur einmal entwickelt werden müssen und dann allen Anwendungen zur Verfügung stehen, während die Migration individuell für jede Anwendung erfolgen muß. Ein weiterer Vorteil dieses Konzepts liegt darin begründet, daß die ursprüngliche Funktionalität der Anwendung erhalten bleibt. Schließlich ist auch der mit dem Emulationskonzept verbundene Zwang, bereits zeitnah für künftige Emulationsfälle technische und organisatorische Vorkehrungen zu treffen und insbesondere die Anwendung in ihren Bezügen zu den Informationen in ihrer Systemumgebung vollständig zu dokumentieren, ein Vorteil für die lang- aber auch mittelfristige Transparenz der Anwendung und damit auch für eine effiziente Wartung.

Im Prinzip und über einige Computergenerationen hinweg wird die Emulation bereits heute mit Erfolg angewandt. So versehen Hersteller von EDV-Systemen ihre neuen Kreationen mit Emulatoren, welche die vorherige Generation nachahmen, damit die Kunden ihre für die früheren Systeme entwickelte spezifische Anwendungssoftware weiterhin in voller Funktionalität betreiben können. IBM stellt für sein System 360 eine Softwareemulation zur Verfügung, um darauf für das frühere System 780 entwickelte Anwendungssoftware laufen zu lassen, Apple stellt für den Power PC eine Emulation bereit, um den Processor Motorala 68000 nachzubilden. Freaks stellen im Internet kostenlos Emulatoren zur Verfügung, damit für frühere und teilweise längst ausgestorbene Hardware-Generationen entwickelte Video- oder Computerspiele auch auf modernen Systemen laufen.

Der archivarischen und übrigens auch der historisch-quellenkritischen Sichtweise kommt der Emulationsgedanke entgegen, was als Vorteil zu werten ist. Da die Emulation ermöglicht, Informationen im jeweils zeitgebundenen Kontext von Anwendung und Funktionalität zugänglich zu machen und zu halten, nicht mehr und nicht weniger, hält sie den Entstehungszusammenhang der digitalen Informationen evident. Während die Migration allenfalls bloße

Daten und Fakten authentisch stabilisiert, ist die Emulation in der Lage, auch die Funktionen rekonstruierbar in der authentischen Systemumgebung zu stabilisieren, die ehemals mit den Informationen verbunden waren. Der intrinsische Wert digitalen Archivguts[15] kann damit nur mit Hilfe der Emulation langfristig rekonstruiert und nachvollziehbar gemacht werden. Mit der Emulation soll die Anwendung in ihrer Funktionalität und ihrem zeitbedingten Zustand einer bedingten Hard- und Softwareumgebung dauerhaft stabilisiert werden. Dies hat aus der Sicht des Archivars mit Authentizität zu tun. Präsentationsmodelle von Archivgut aus den späten 90er Jahren unseres Jahrhunderts oder gar Dokumente aus Büroautomationssystemen unserer Zeit in die zweifellos geänderte Funktionalität der Zukunft zu migrieren, vermittelt von vornherein ein schiefes Bild von dem, was Jahrzehnte zuvor möglich war.

Im Gegensatz zur Migration erfordert das Emulationskonzept zeitnahe Vorkehrungen. Für eine spätere Emulation eines früheren Zustands ist es notwendig, die digitalen Objekte, die Metadaten und Kontextinformationen zusammen mit der Anwendungssoftware und dem Betriebssystem, das notwendig ist, um die Anwendungssoftware zum Laufen zu bringen sozusagen zusammen mit der Dokumentation zu konservieren, welche das Zusammenspiel der Komponenten darstellt und die Bedienung erläutert. Mit den Methoden der Datenauffrischung und der Formatanpassung wird dafür gesorgt, daß die maschinenlesbaren Teile des Pakets lesbar bleiben. Bei Bedarf soll künftig durch Emulation der früheren Betriebssystemebene dafür gesorgt werden, daß das Anwendungsprogramm läuft, das seinerseits den Zugriff und die Darstellung der digitalen Objekte und der dazu gehörenden Daten sicherstellt.

Auch das Emulationskonzept wird bei Anwendungen mit Dokumenten, deren Darstellung und deren Funktionalität in den Dokumenten selbst mit SGML oder Derivaten davon maschinen- und menschenlesbar beschrieben ist, einfacher und kostengünstiger zu realisieren sein, als bei komplexen Anwendungen, wie sie bei den Dokumenten-Management-Systemen dargestellt wurden. Für die Langzeitspeicherung und Langzeitverfügbarkeit retrodigitalisierten Archiv- oder Bibliotheksguts hat dies die Konsequenz, daß neben dem hybriden Systemansatz die Kapazität und die funktionale Leistungsfähigkeit von unter Verwendung standardisierter Beschreibungssprachen erstellten Datei-Systemen eingehender untersucht werden müssen, in welche die digitalen Objekte eingebettet sind. In diesem Zusammenhang sind insbesondere zwei wichtige Meilensteine zu erreichen. Zum einen die rationelle Generierung von HTML-Anwendungen mit eingebetteten Objekten mit Hilfe von

---

15 Vgl. Angelika *Menne-Haritz* und Nils *Brübach:* Der intrinsische Wert von Archiv- und Bibliotheksgut. Kriterienkatalog zur bildlichen und textlichen Konversion bei der Bestandserhaltung. Marburg 1997. Insbes. S. 29ff.

Standardwerkzeugen. Zum anderen beispielhafte Präsentationsmodelle, welche die erweiterten Funktionalitäten des künftigen Standards XML nutzen. Beide Vorhaben verfolgen das Ziel, nicht in komplexen Systemumgebungen, sondern in standardisiert beschriebenen Dokumenten digitale Objekte einzubetten, zusammen mit Metadaten, Kontextinformationen und Funktionalitäten, nicht zuletzt, da ihre Langzeitverfügbarkeit dann nicht nur von Maschinen abhängig ist, sondern solche integrierten Dokumente notfalls sogar mit dem menschlichen Auge zu lesen und mit dem menschlichen Verstand zu interpretieren sind.

Verzeichnis der Autorinnen und Autoren

Prof. Dr. Gerhard Banik
Staatliche Akademie der Bildenden Künste Stuttgart, Studiengang Restaurierung und Konservierung von Graphik, Archiv- und Bibliotheksgut
Höhenstraße 16, 70736 Fellbach

Dr. Jürgen Bunzel
Deutsche Forschungsgemeinschaft, Referat Wissenschaftliches Bibliothekswesen
53170 Bonn

Dr. Marianne Dörr
Bayerische Staatsbibliothek
Ludwigstraße 16, 80539 München

Dr. Hans-Heinrich Ebeling
Stadtarchiv Duderstadt
Christian-Blank-Str.1, 37115 Duderstadt

Dr. Peter Exner
Landesarchivdirektion Baden-Württemberg
Abt. Landesforschung und Landesbeschreibung, Außenstelle Karlsruhe
Nördliche Hildapromenade 2, 76133 Karlsruhe

Dr. Thomas Fricke
Landesarchivdirektion Baden-Württemberg
Eugenstraße 7, 70182 Stuttgart

Dr. John W. Haeger
Research Libraries Group
1200 Villa Street, Mountain View, Ca 94041–1100, USA

Dr. Hermann Leskien
Bayerische Staatsbibliothek
Ludwigstraße 16, 80539 München

Prof. Dr. Sönke Lorenz
Institut für geschichtliche Landeskunde und historische Hilfswissenschaften
Universität Tübingen
Wilhelmstraße 36, 72074 Tübingen

Dr. Norbert Lossau
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Platz der Göttinger Sieben 1, 37073 Göttingen

Dr. Gerald Maier
Landesarchivdirektion Baden-Württemberg
Eugenstraße 7, 70182 Stuttgart

Dr. Carol Mandel
New York University, 11th floor, Bobst Library
70 Washington Square West, New York, N.Y. 10012, USA

Dr. Deanna B. Marcum
Council on Library and Information Resources
1755 Massachusetts Avenue, N.W., Suite 500, Washington,
D.C. 20036–2188, USA

PD Dr. Angelika Menne-Haritz
Archivschule Marburg
Bismarckstraße 32, 35037 Marburg

Prof. Dr. Elmar Mittler
Niedersächsische Staats- und Universitätsbibliothek Göttingen
Platz der Göttinger Sieben 1, 37073 Göttingen

Hans Rütimann
Council on Library and Information Resources
312 West 77th Street, Suite G, New York, N.Y. 10024–6883, USA

Dr. Karsten Uhde
Archivschule Marburg
Bismarckstraße 32, 35037 Marburg

Prof. Dr. Hartmut Weber
Bundesarchiv
Potsdamer Str. 1, 56075 Koblenz

Prof. August Wegmann
Fachhochschule Karlsruhe – Hochschule für Technik, Fachbereich
Wirtschaftsinformatik
Moltkestraße 30, 76133 Karlsruhe

ISBN 3-17-016062-1