

Grundlagen der Retrodigitalisierung von Texten und Bildern

Referat im Rahmen der Lehrveranstaltung

Wissenschaftliches Bibliothekswesen in Deutschland

(Lehrkörper: B. Dugall)

an der Bibliotheksschule in Frankfurt am Main,

Kursjahr 2000/2001,

gehalten am 16.02.2001.

Referent: **Dr. Jan Harloff**

Übersicht:

1.	Einleitung _____	2
2.	Digitale Erfassung _____	3
	2.1. Direkte Digitalisierung	3
	2.2. Verfilmung und Digitalisierung	4
3.	Dateiformate für die Retrodigitalisierung _____	4
	3.1. Digitaler Master	5
	3.2. Image-Version	5
	3.3. Download-Version	6
4.	Erfassung retrodigitalisierter Texte _____	7
	4.1. Manuelle Texterfassung	7
	4.2. Automatisierte Texterfassung	7
	4.3. Strukturbeschreibung	8
5.	Adressierung der retrodigitalisierten Dateien im Internet _____	9
6.	Langzeitarchivierung _____	10
7.	Urheberrecht _____	10
8.	Nutzen der Retrodigitalisierung _____	11
9.	Problem der Titelauswahl _____	12
10.	Literatur _____	13

1. Einleitung:

Seit mehreren Jahren laufen, vor allem in den USA, aber auch an den großen Nationalbibliotheken in Europa, wie der British Library oder der Bibliothèque Nationale, Vorhaben im Bereich Retrodigitalisierung (ENDERLE 1997). Hieran knüpfte die DFG (Deutsche Forschungsgemeinschaft) an, indem sie 1997 im Förderbereich "Verteilte Digitale Forschungsbibliothek" das Programm "Retrospektive Digitalisierung von Bibliotheksbeständen" gestartet hat (DFG 1997, 1998). Zur Vorbereitung des neuen Förderprogramms hat die DFG zwei Arbeitsgruppen zusammenberufen, eine für inhaltliche und eine für technische Fragen. Die Ergebnisse beider Arbeitsgruppen sind in einem gemeinsamen Band veröffentlicht worden und bieten einen hervorragenden Überblick über die hier angesprochene Thematik, können allerdings natürlich nicht mehr auf dem neuesten Stand sein (AG INHALT 1997; AG TECHNIK 1997).

Der Begriff "Digitalisierung" bezeichnet die Überführung eines analog vorliegenden Dokuments, z.B. gedruckter Literatur, in eine digitale Datei, die mit den Methoden der EDV gehandhabt werden kann. Man spricht von einer "elektronischen Sekundärform", die aus der "Primärform" generiert wird. Der Zusatz "retrospektiv" deutet an, daß die Publikation der Primärform zeitlich zurück liegt und die Erstellung der elektronischen Version nicht zeitgleich vorgenommen wurde, was man im Gegensatz dazu als "elektronische Parallelfom" bezeichnet (LESKIEN 2000).

Die Motivation für die Durchführung von Retrodigitalisierungsprojekten hat grundsätzlich zwei Aspekte (SITTS 2000):

- 1.) Die Bestandserhaltung, also den Schutz der Primärform und die Sicherung ihres Inhalts durch Langzeitarchivierung der elektronischen Sekundärform.
- 2.) Die Benutzung, also die Verfügbarmachung des Inhalts der Primärform durch Bereitstellung der elektronischen Sekundärform im Internet.

Der Punkt der Langzeitarchivierung der elektronischen Sekundärform als Motivation wird inzwischen allerdings durchaus kritisch gesehen, da die Frage der Sicherstellung einer dauerhaften Verfügbarkeit elektronischer Dateien sich als sehr schwierig erwiesen hat (LESKIEN 2000). Auf diesen Punkt wird in Abschnitt 6 noch näher eingegangen.

2. Digitale Erfassung:

Der erste Schritt der Digitalisierung einer Primärform ist die Erfassung der Vorlage mittels eines Scanners. Dies kann grundsätzlich auf zwei verschiedenen Wegen erfolgen (AG TECHNIK 1997):

- 1.) Die Digitalisierung direkt von der Vorlage.
- 2.) Die Verfilmung der Vorlage und anschließende Digitalisierung vom Mikrofilm.

2.1. Direkte Digitalisierung:

Für eine direkte Digitalisierung von der Vorlage sind die neuen Typen der Buch- oder Aufsichtsscanner besonders geeignet, vor allem beim Umgang mit historisch wertvollen und besonders schonungsbedürftigen Beständen. Bei diesen Scannertypen ist der Lesekopf oberhalb der Vorlage angebracht und die eigentliche Abtastung erfolgt berührungslos (AG TECHNIK 1997). Die Software des Scanners erzeugt eine Bilddatei, in der das Bild als Rastergraphik, also in Form von einzelnen Pixeln, dargestellt wird. Für die Erstellung dieser Rastergraphiken müssen einige grundsätzliche Entscheidungen gefällt werden, vor allem bzgl. der Auflösung und der Farbtiefe der zu erzeugenden Bilder, da diese Faktoren entscheidend den späteren Speicherbedarf und die weitere Verwendbarkeit der Bilddateien beeinflussen (PUGLIA 2000).

Die sinnvollerweise zu wählende Auflösung hängt ab von der Vorlage und der geplanten Verwendung der Bilddateien. Bei gedruckten Texten mit gängigen Schriftgrößen wird eine Auflösung von wenigstens 300 dpi (dots per inch) empfohlen. Falls eine spätere Texterkennung geplant ist, dann sollte die Auflösung mindestens 400 dpi betragen. Wenn es sich bei der Vorlage um eine Photographie handelt, dann können höhere Auflösungen sinnvoll sein, je nach Detailreichtum bis zu 600 dpi (AG TECHNIK 1997).

Für die Farbtiefe sind die folgenden Formate gängig (PUGLIA 2000):

- 1.) 1-Bit Schwarzweiß: bitonal.
- 2.) 8-Bit Graustufen: 256 Graustufen.
- 3.) 8-Bit Farbe: 256 Farben.
- 4.) 24-Bit Farbe: ca. 17 Mio. Farben.

Wenn es sich bei der Vorlage um einen gedruckten Text handelt, dann wird häufig eine bitonale Erfassung mit einem Bit pro Pixel ausreichen. Bei schwierigeren schwarzweißen

Vorlagen, etwa Handschriften oder Bleistiftzeichnungen, sollte eher mit 256 Graustufen digitalisiert werden (AG TECHNIK 1997). Während bei simplen farbigen Vorlagen 8-Bit Farbe durchaus ausreichen kann, wird bei Photographien meist 24-Bit Farbe sinnvoll sein (PUGLIA 2000). Grundsätzlich ist noch zu erwähnen, daß beim Digitalisieren mit Graustufen die Auflösung, verglichen mit einer bitonalen Version, bei gleicher Wiedergabequalität reduziert werden kann. Bei Verwendung von 256 Graustufen kann so für gedruckte Texten mit gängigen Schriftgrößen eine Auflösung von 250 dpi durchaus ausreichen (DÖRR & WEBER 1997).

2.2. Verfilmung und Digitalisierung:

Die oben erwähnte Unsicherheit bzgl. der langfristigen Verfügbarkeit elektronischer Dateien hat dazu geführt, daß zunehmend empfohlen wird, die Digitalisierung nicht mehr an den Vorlagen selbst vorzunehmen. Statt dessen wird als neues Verfahren vorgeschlagen, die Vorlagen zunächst zu verfilmen, und die Digitalisierung anschließend von den erzeugten Mikrofilmen vorzunehmen, mit Hilfe von Mikrofilmscannern. Tatsächlich ist die Digitalisierung von Mikrofilmen besonders wirtschaftlich und führt zu sehr guten Ergebnissen (DÖRR & WEBER 1997). Bei der Auswahl des Mikrofilms sollte allerdings auf geeignetes Filmmaterial auf Polyestergrundlage geachtet werden, das als alterungsbeständiger Informationsträger gilt. Mit diesem Verfahren erzeugt man dann sowohl einen analogen Langzeitspeicher für die Archivierung, als auch eine elektronische Sekundärform für die Benutzung (AG TECHNIK 1997). Für ein solches Vorgehen, oder sogar für Hybridsysteme, die Digitalisierung und Verfilmung in einem Arbeitsschritt durchführen, entscheiden sich zunehmend mehr Stellen, wie z.B. die österreichische Nationalbibliothek (SMA 1999).

3. Dateiformate für die Retrodigitalisierung:

Nachdem die Bilddateien durch Digitalisierung erzeugt wurden, ist die Bandbreite von möglichen Formaten für deren Speicherung enorm. Für die Frage der Wahl des Formats ist der geplante weitere Einsatz der Dateien von entscheidender Bedeutung. Man kann die Einsatzbereiche grundsätzlich unterscheiden in (AG TECHNIK 1997):

- 1.) Datei für die Langfristarchivierung und als digitaler Master für die Erstellung weiterer Dateiformate.
- 2.) Datei als Image-Version für die direkte Benutzung am Bildschirm.
- 3.) Datei als Download-Version für das Herunterladen durch Benutzer.

3.1. Digitaler Master:

Für den digitalen Master hat sich das Format TIFF (Tagged Image File Format) der Firma Aldus als gängigste Variante durchgesetzt. Besonders günstig für bibliothekarische Zwecke ist der Umstand, daß jeder Bilddatei Informationen beigegeben werden können, die in ein zugehöriges "Image File Directory" geschrieben werden. Hier kann man etwa Aussagen speichern über technische Details der Digitalisierung, wie Auflösung, Farbtiefe, usw. Wegen der großen Speichermenge, eine bitonale DIN A4 Seite mit 400 dpi Auflösung benötigt ca. 2 MB Speicherplatz, eignet sich allerdings das unkomprimierte TIFF nicht. Empfohlen wird daher die Verwendung der verlustfreien Kompression TIFF G 4, bei der sich der Speicherbedarf des genannten Beispiels auf ca. 100 bis 150 KB reduziert (AG TECHNIK 1997).

3.2. Image-Version:

Für die Bereitstellung der Bilddateien über das WWW (World Wide Web) sind die relevantesten Faktoren vor allem eine Unterstützung des Formats durch gängige Web-Browser und die Größe der Datei, um einen schnellen Datentransfer und eine gute Performanz zu gewährleisten. Hierfür gibt es vor allem zwei gängige Formate: GIF und JPEG (AG TECHNIK 1997).

Das Format GIF (Graphics Interchange Format) wurde 1987 von der Firma CompuServe für Speicherung und Austausch von Rastergraphiken entwickelt und ist seitdem eines der weitverbreitetsten Graphikformate im Internet geworden (MCMURDO 1997). Es ist vor allem attraktiv wegen seiner hardwareübergreifenden Verbreitung als Austauschformat und der Verwendung des verlustfreien Kompressionsverfahrens LZW. Da GIF jedoch lediglich eine Farbtiefe bis zu 8 Bit zuläßt, eignet es sich nur für bitonale und Halbtonvorlagen (AG TECHNIK 1997).

Der Name des Formats JPEG leitet sich von dem Komitee ab, das es erstellt hat, der Joint Photographic Experts Group. Es handelt sich hierbei genau genommen weniger um ein Format als vielmehr um eine ganze Familie sehr effektiver Kompressionsalgorithmen für die Komprimierung digitaler Standbilder (MCMURDO 1997). Der Kompressionsfaktor ist bei der Speicherung vom Anwender frei wählbar, wobei allerdings zu bedenken ist, daß das hier verwendete Kompressionsverfahren immer mit einem Informationsverlust einhergeht, der sich mit zunehmendem Kompressionsfaktor stärker bemerkbar macht (AG TECHNIK 1997). Da die Speicherung unter JPEG immer mit einer Farbtiefe von 24 Bit geschieht, lohnt sich

der Einsatz dieses Formats nur für sehr anspruchsvolle Farb- oder Graustufenbilder (MCMURDO 1997).

Im Jahre 1996 wurde von einer Gruppe von Graphik- und Programmierungsspezialisten das neue Format PNG (Portable Network Graphics) entwickelt, und das W3C (World Wide Web Consortium) bemühte sich sehr um eine Verbreitung dieses neuen Formats (MEISSNER 1996; MCMURDO 1997). In ihrem Bericht von 1997 hat auch die AG Technik PNG als ein vielversprechendes neues Dateiformat für den Bereich Retrodigitalisierung empfohlen, sowohl was die Verwendung als digitaler Master angeht, als auch für die Bereitstellung über das WWW. Dieses Format bietet tatsächlich einige Vorteile gegenüber TIFF und GIF, sowohl aus lizenzrechtlichen als auch technischen Gründen. Ein Problem dieser beiden Formate liegt darin, daß die Unisys Corporation vor einigen Jahren das Patentrecht für das gängige LZW-Komprimierungsverfahren erworben hat, auf dem die Komprimierung von GIF basiert und im Falle von farbigen Bildern auch von TIFF. In der Folge hat Unisys Lizenzgebühren von den Anbietern gefordert, die ihre Bilddateien im kommerziellen Bereich einsetzen. In Zukunft sind in diesem Bereich auch für Bibliotheken Probleme oder zusätzliche Kosten nicht auszuschließen. Im Gegensatz dazu ist die von PNG eingesetzte DEFLATE-Komprimierung nicht nur lizenzfrei, sondern auch um ca. 10 bis 30% effektiver als LZW (AG TECHNIK 1997). Nichtsdestotrotz hat sich PNG, zumindest im deutschen Bibliotheksbereich, nicht durchsetzen können. Praktisch durchgängig hat es sich hierzulande eingebürgert, den digitalen Master in TIFF zu speichern und die Dateien über das WWW entweder mittels GIF zur Verfügung zu stellen (siehe z.B. DÖRR 2000), oder aber mittels JPEG, wenn es sich um anspruchsvollere Graphiken handelt (siehe z.B. GEIBELMANN & KECKS 2000).

3.3. Download-Version:

Das Herunterladen kompletter retrodigitalisierter Dateien ist für eine bequeme Benutzbarkeit sehr relevant, vor allem im Hinblick auf potentiell langsame Übertragungsgeschwindigkeiten im Internet und eine schlechte Performanz beim Arbeiten mit der Datei über das WWW. Die einmal lokal gespeicherte Version ist dann für den Benutzer die Grundlage für Bildschirm- oder Druckerausgabe. Hier bieten sich Formate an, die speziell für die Beschreibung des Layouts ganzer Dokumente entwickelt wurden, in erster Linie zu nennen ist das Produkt PDF (Portable Document Format) der Firma Adobe. Der Vorteil von PDF ist, daß das Dokumentlayout hardwareunabhängig festgelegt ist und für eine Darstellung auf Bildschirm und Drucker gleichermaßen geeignet ist (AG TECHNIK 1997).

4. Erfassung retrodigitalisierter Texte:

In den meisten Fällen der bibliothekarisch relevanten Retrodigitalisierungsprojekte handelt es sich bei den Vorlagen um Textdokumente, also z.B. um Bücher. In diesem Falle sollte das Verfahren sich nicht auf die oben beschriebene Erstellung von Rastergraphiken beschränken, vielmehr sollte als weiterer Schritt eine Texterfassung erfolgen. Hierdurch wird eine Navigation im Dokument ermöglicht, die sich ansonsten, bei sequentiell durchblättern des Texts am Bildschirm, äußerst mühselig gestaltet (AG INHALT 1997). Man spricht hier vom Mehrwert der elektronischen Sekundärform, da Optionen ermöglicht werden, die mit der Primärform so nicht durchführbar sind, wie etwa die gezielte Suche nach Wörtern im Text (LESKIEN 2000). Grundsätzlich gibt es für eine Texterfassung drei mögliche Stufen (AG TECHNIK 1997):

- 1.) Die manuelle Erfassung der Texte.
- 2.) Die automatisierte Erfassung durch Texterkennungsprogramme.
- 3.) Die Erstellung einer Strukturbeschreibung.

4.1. Manuelle Texterfassung:

Eine manuelle Volltexterfassung, d.h. das Abtippen des kompletten Textes, ist wegen der damit verbundenen hohen Personalkosten kaum denkbar. Es wird aber empfohlen, einzelne Strukturelemente eines Buches manuell zu erfassen und zu indizieren, wobei als Mindeststandard das Inhaltsverzeichnis gilt und zusätzlich das Register sehr empfohlen wird. Durch eine Verlinkung der so erstellten Indizes mit den zugehörigen Textseiten wird ein gezielter inhaltlicher Zugriff auf einzelne Teile des Textes ermöglicht. Ein extremer Mehrwert könnte hier erreicht werden, wenn die einzelnen Register der digitalisierten Bücher zusammengespielt werden, z.B. innerhalb einer retrodigitalisierten Sammlung, oder aber sogar sammlungsübergreifend. Ein derart kumulierter Index würde dann den Zugriff von einem bestimmten Begriff auf alle Bücher erlaubt, in denen dieser Begriff vorkommt (AG TECHNIK 1997).

4.2. Automatisierte Texterfassung:

Für die Volltexterfassung in Rastergraphiken bietet sich die automatisierte Erkennung von Texten an, ein Verfahren, daß man als OCR (Optical Character Recognition) bezeichnet. Es gibt eine Vielzahl von OCR-Programmen, die sich entweder des Prinzips des Muster

vergleichs oder aber der Merkmalanalyse bedienen, um in einer Rastergraphik Zeichen zu erkennen und diese in ASCII (American Standard Code for Information Interchange) umzusetzen. Bei sauberen Vorlagen mit leicht lesbaren Schriften in guter Druckqualität lassen sich mit OCR inzwischen Erkennungsquoten von über 99% erreichen. Verschmutzungen, uneinheitlicher Schriftsatz, mangelhafte Schriftqualität und veraltete Schriftarten, wie etwa Fraktur, bereiten OCR-Programmen jedoch ernsthafte Probleme. Verschiedene Tests mit Texten aus dem 19. Jahrhundert haben ergeben, daß hier lediglich Erkennungsquoten von 60 bis 70% zu erwarten sind (AG TECHNIK 1997). Es erscheint daher häufig sinnvoll, die erzeugte ASCII-Version lediglich für eine Volltextsuche im Hintergrund bereit zu halten, während für die eigentliche Darstellung trotzdem die Rastergraphik verwendet wird. Dies hat zusätzlich den Vorteil, daß hierdurch der Charakter des Textoriginals und seine formale Gestaltung erhalten bleiben, die bei der Textkonversion verloren gehen (LESKIEN 2000).

4.3. Strukturbeschreibung:

An eine automatisierte Volltexterfassung kann sich die Erstellung einer Strukturbeschreibung der Vorlage anschließen, wobei allerdings der hierfür nötige Personalaufwand enorm ist. Unter der Strukturbeschreibung eines Dokuments versteht man die Kennzeichnung distinktiver struktureller Elemente des Textes, wie etwa Überschriften, Absätze, usw. Anders als bei dem in Abschnitt 3.3 erwähnten Format PDF geht es hierbei also weniger um das Layout eines Dokuments, als vielmehr um dessen logische Strukturierung. Als Werkzeug hat man hierfür die sogenannten Beschreibungssprachen, von denen einige relevante auf dem übergeordneten Dokumentenbeschreibungsformat SGML (Standard Generalized Markup Language) aufbauen, das in der ISO-Norm 8879 von 1986 festgelegt ist. Ein SGML-Dokument besteht dabei ganz grundsätzlich aus drei Teilen (AG TECHNIK 1997):

- 1.) Vereinbarung der Syntax (SGML-Deklaration).
- 2.) Definition des Dokumenttyps, DTD (Document Type Definition).
- 3.) Ausprägung des Dokuments (das eigentliche Dokument selbst).

Eine der bekanntesten Weiterentwicklungen von SGML ist HTML (Hypertext Markup Language), auf der die Darstellung von Web-Sites im WWW basiert. Für die Strukturbeschreibung retrodigitalisierter Textdokumente eignet sich aber vor allem eine andere Weiterentwicklung von SGML, und zwar die sogenannte XML (eXtensible Markup Language), die 1998 vom W3C als neuer Web-Standard festgesetzt wurde (TRENDS & MÄRKTE 1999). In seinen Strukturen ist XML teilweise der HTML verwandt, geht aber in

den Funktionalitäten weit darüber hinaus. Einer der wichtigsten Unterschiede ist, daß die im XML verwendeten Befehle, die sogenannten Tags, keine Information darüber enthalten, wie sie dargestellt werden sollen. Dies wird statt dessen in einem der eigentlichen Datei vorgeschalteten Kopf vereinbart. Durch diese Struktur ist XML sehr gut erweiterbar und anpaßbar für spezielle Aufgaben (VAN DER WERF 1998).

XML ist dabei, sich zu einem grundsätzlichen Standard für die Bereitstellung elektronischer Ressourcen über das Internet zu entwickeln. Es ist daher zu erwarten, daß immer mehr elektronische Primär- und Parallelformen in XML publiziert werden (TRENDS & MÄRKTE 1999). Zusätzlich wird derzeit die Verwendbarkeit von XML für Zwecke der elektronischen Archivierung recht positiv bewertet. Es bietet sich daher auch für Bibliotheken an, XML für die strukturierte Archivierung und Bereitstellung retrodigitalisierter Sekundärformen zu verwenden, in einer Form, die kompatibel zu anderen elektronischen Ressourcen ist (VAN DER WERF 1998).

5. Adressierung der retrodigitalisierten Dateien im Internet:

Damit die zur Verfügung gestellten Sekundärformen verwendbar sind, muß der Benutzer sie identifizieren können. Dies geschieht anhand zugeordneter Namen, die gemäß eines definierten Benennungsschemas gebildet werden. Man unterscheidet ortsgebundene und ortstransparente Namen. Ortsgebundene Namen identifizieren einen Ort und damit indirekt die Ressource, die sich an diesem Ort befindet. Ortstransparente Namen bezeichnen eine Ressource selbst. Um auf die so bezeichnete Ressource zugreifen zu können, muß der ortstransparente Name in eine Ortsangabe übersetzt werden, dies geschieht üblicherweise durch sogenannte Name Server. Das übliche Schema zur Benennung von Ressourcen im Internet ist der URL (Uniform Resource Locator). Der größte Nachteil eines URL ist allerdings seine mangelnde Persistenz, bereits nach wenigen Monaten sind viele URLs nicht mehr gültig. Die IETF (Internet Engineering Task Force) versucht daher seit einigen Jahren, ein eindeutiges, persistentes und ortstransparentes Benennungsschema für Internet-Ressourcen zu entwickeln, die URNs (Uniform Resource Names). Eindeutig und persistent bedeutet, daß jeder Name nur einmal vergeben und dauerhaft mit einer Ressource verknüpft wird. Diese Verknüpfung bleibt erhalten, sogar wenn die Ressource selbst nicht mehr existiert oder nicht mehr zugreifbar ist. Solche Bemühungen sind für die langfristige Perspektive der Verfügbarmachung retrodigitalisierter Dateien von größter Bedeutung (AG TECHNIK 1997).

6. Langzeitarchivierung:

Wie eingangs bereits erwähnt, wird inzwischen die Retrodigitalisierung als Mittel der Informationssicherung über Langzeitarchivierung der elektronischen Sekundärform sehr kritisch bewertet (LESKIEN 2000). Das Hauptproblem entsteht dabei ironischerweise durch einen Umstand, der ansonsten durchaus positive Folgen hat, nämlich die extrem schnelle Innovation auf dem EDV-Sektor, sowohl im Hardware- wie im Softwarebereich. Hierdurch besteht die große Gefahr, daß verhältnismäßig schnell die notwendigen Geräte, Technologien und Formate, die zum Lesen der gespeicherten Sekundärformen notwendig sind, nicht mehr zur Verfügung stehen (REINITZER & KROLLER 1996). Zur Zeit gibt es vor allem zwei Modelle, wie man diesem Problem begegnen könnte (MEYENBURG 2000):

- 1.) Eine stetige Migration der Datenbestände, um sie an die neuere Hard- und Software anzupassen.
- 2.) Eine stetige Entwicklung von Emulationen älterer Hard- und Software, um auf der Basis modernster Technik die jeweils dokumentadäquate, veraltete Umgebung zu erzeugen.

Solange nicht eins der beiden Modelle seine Tauglichkeit in der Praxis unzweifelhaft erwiesen hat, gehen die Empfehlungen dahin, bei einer erwünschten Langzeitarchivierung den oben beschriebenen Weg der Kombination aus Verfilmung und Digitalisierung zu wählen, in dem der Mikrofilm die Rolle des alterungsbeständigen Informationsträgers übernimmt (DÖRR & WEBER 1997; LESKIEN 2000).

7. Urheberrecht:

Das deutsche UrhG (Urheberrechtsgesetz) schützt den Urheber eines Werkes, indem es ihm sogenannte Ausschließlichkeitsrechte einräumt. Vor allem darf danach ausschließlich der Urheber sein Werk vervielfältigen. Eine Vervielfältigung ist im Sinne des UrhG auch die Übertragung eines Werkes auf einen anderen Träger, womit es sich bei der Digitalisierung ohne Zweifel um einen Akt der Vervielfältigung handelt. Andere Personen oder Institutionen, z.B. Bibliotheken, dürfen das Werk, abgesehen von einigen Ausnahmefällen, nur dann vervielfältigen, wenn der Urheber dies erlaubt. Das Urheberrecht ist allerdings ein zeitlich begrenztes Recht, das 70 Jahre nach dem Tod des Urhebers erlischt (PETERS 1998). Dies gilt im Grunde genauso im amerikanischen Copyright (SMITH LEVINE 2000). Um hier Problemen aus dem Weg zu gehen, liegt der Schwerpunkt deutscher Retrodigitalisierungsprojekte daher auf alten, rechtsfreien Dokumenten (LESKIEN 2000).

Dieser Zustand kann auf Dauer nicht sinnvoll sein, ein Schwerpunkt zukünftiger Bemühungen der Bibliotheken muß es daher sein, in Kooperation mit Verlagen und sonstigen Inhabern von Urheberrechten Lösungen zu erarbeiten, so daß die Retrodigitalisierung auch auf urheberrechtsrelevante Literatur ausgedehnt werden kann (AG INHALT 1997). Einer der möglichen Ansatzpunkte für eine solche Lösung liegt in dem von mehreren Großverlagen und Verlegerverbänden auf der Frankfurter Buchmesse 1997 vorgeschlagenen System DOI (Document Object Identifier). Das DOI-System verfolgt grundsätzlich zwei Ziele (GABEL 1998):

- 1.) Es soll eine persistente Identifizierung von Verlagsprodukten im Internet gewährleistet werden. Hierbei handelt es sich also um eine ähnliche Funktion wie bei den in Abschnitt 5 beschriebenen URNs.
- 2.) Es soll eine Basis für die Abrechnung bei Transaktionen mit den Verlagsprodukten im Internet geschaffen werden.

Als Zentrum des Verfahrens fungiert das DOI-Directory, das für die Registrierung der Firmen wie der Produkte sorgt. Das DOI-Directory agiert dabei als Verteiler und Mittler zwischen dem Nutzer und dem Anbieter. Wenn ein Nutzer eine DOI anwählt, dann wird eine Meldung an das DOI-Directory gesandt, das wiederum eine Antwort schickt mit der aktuellen Web-Adresse des Anbieters, die den Nutzer dahin weiterleitet (GABEL 1998).

8. Nutzen der Retrodigitalisierung:

Die hier dargestellten Ausführungen machen deutlich, was für ein großes Potential in der Retrodigitalisierung liegt. Zusammengefaßt kann man sagen, daß es vor allem drei große Bereiche gibt, in denen die Stärken der Retrodigitalisierung liegen (CONWAY 2000):

- 1.) Die Primärform wird geschützt. Durch die Verfügbarmachung der elektronischen Sekundärform wird ein Großteil der Benutzung der Primärform unnötig, so daß diese vor Beschädigungen geschützt wird.
- 2.) Der Inhalt der Primärform wird zugänglich. Durch die Möglichkeit des zeit- und ortsunabhängigen Zugriffs auf die elektronische Sekundärform über das Internet sind schlecht zugängliche oder stark genutzte Ressourcen für den Benutzer bequem verfügbar.
- 3.) Im Vergleich zu der Primärform wird ein Mehrwert erzielt. Vor allem bei Sekundärformen, die hochwertig im Volltext oder sogar strukturell erfaßt sind, ergibt

sich ein deutlicher Mehrwert, da Nutzungen ermöglicht werden, die mit der Primärform nicht realisierbar wären.

9. Problem der Titelauswahl:

Bei den überzeugenden Vorteilen der Retrodigitalisierung stellt sich die Frage, ob es nicht erstrebenswert wäre, den gesamten deutschen Bibliotheksbestand zu digitalisieren. Durch eine simple Kalkulation läßt sich allerdings zeigen, daß dieser Ansatz komplett unrealistisch ist. Das Problem bei der Retrodigitalisierung ist der große Personalaufwand und die sich daraus ergebenden Kosten bei der Durchführung. Berechnungen des Aufwands belaufen sich für eine Retrodigitalisierung mit manueller Texterfassung nur des Inhaltsverzeichnisses auf Kosten von ca. 0,70 bis 1,00 DM und einen Zeitbedarf von ca. 6 Sekunden pro Seite. Bei einer geschätzten Menge von 13 Millionen Titeln an deutschen Hochschulbibliotheken mit einer durchschnittlichen Seitenzahl von 400 Seiten pro Buch ergibt sich daraus eine untere Grenze des Finanzaufwands von 3,64 Mrd. DM bei einem Zeitaufwand von 5500 Mannjahren (BENZ 1998).

Darüber hinaus lassen sich keinesfalls Tendenzen für wesentliche Kostensenkungen ausmachen, da Bücher individuell verfaßt und gestaltet sind. Sie entziehen sich daher einer rein maschinellen Digitalisierung, und Personalkosten werden in Zukunft eher steigen als sinken. Die komplett digitale deutsche Hochschulbibliothek im Sinne einer elektronischen Verfügbarkeit aller vorhandenen Bestände ist also aus heutiger Sicht eine Utopie. Damit ergibt sich die Frage der Auswahl, die Entscheidung, welcher Teil des Bestandes sinnvollerweise digitalisiert werden soll, als das zentrale Problem bibliothekarischer Retrodigitalisierungsprojekte (LESKIEN 2000).

10. Literatur:

AG INHALT (1997): "Empfehlungen zur inhaltlichen Auswahl von Bibliotheksmaterialien für die retrospektive Digitalisierung : Bericht der Facharbeitsgruppe Inhalt" ; *Retrospektive Digitalisierung von Bibliotheksbeständen : Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppen 'Inhalt' und 'Technik' (DBI-Materialien ; 166)* ; Mittler E. [Hrsg.] ; Deutsches Bibliotheksinstitut ; Berlin.

Der Text ist zusätzlich als HTML-Datei verfügbar unter:

http://www.sub.uni-goettingen.de/ebene_2/vdf/empfehl.htm

AG TECHNIK (1997): "Die Retrodigitalisierung von Bibliotheksbeständen für eine Verteilte Digitale Forschungsbibliothek : Bericht der Arbeitsgruppe Technik" ; *Retrospektive Digitalisierung von Bibliotheksbeständen : Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppen 'Inhalt' und 'Technik' (DBI-Materialien ; 166)* ; Mittler E. [Hrsg.] ; Deutsches Bibliotheksinstitut ; Berlin.

Der Text ist zusätzlich als HTML-Datei verfügbar unter:

http://www.sub.uni-goettingen.de/ebene_2/vdf/einstieg.htm

BENZ C. (1998): "Die deutsche Hochschulbibliothek auf dem Weg zur Electronic Library" ; *De Officio Bibliothecarii : Beiträge zur Bibliothekspraxis - Hans Limburg zum 65. Geburtstag gewidmet* ; S. 25-39 ; GABEL G., SCHMITZ W., STÄGLICH D. [Hrsg.] ; Greven Verlag ; Köln.

CONWAY P. (2000): "Chapter II : Overview - Rationale for Digitization and Preservation" ; *Handbook for Digital Projects : A Management Tool for Preservation and Access* ; SITTS M.K. [Hrsg.] ; Northeast Document Conservation Center ; Andover, Massachusetts.

Der Text ist zusätzlich als HTML-Datei verfügbar unter:

<http://www.nedcc.org/digital/II.htm>

DFG - Deutsche Forschungsgemeinschaft (1997): *Merkblatt : Retrospektive Digitalisierung von Bibliotheksbeständen* ; Text ist als HTML-Datei verfügbar unter:

http://www.dfg.de/foerder/formulare/1_52.htm

DFG - Deutsche Forschungsgemeinschaft (1998): "Weiterentwicklung der überregionalen Literaturversorgung : Memorandum" ; *ZfBB - Zeitschrift für Bibliothekswesen und Bibliographie* ; Jg. 45 ; Nr. 2 ; S. 135-164.

Der Text ist zusätzlich als HTML-Datei verfügbar unter:

<http://www.dfg.de/foerder/biblio/memo.html>

DÖRR M. (2000): "Das Münchener Digitalisierungszentrum an der Bayerischen Staatsbibliothek" ; *BFB - Bibliotheksforum Bayern* ; Jg. 28 ; Nr. 2 ; S. 154-164.

DÖRR M., WEBER H. (1997): "Digitalisierung als Mittel der Bestandserhaltung : Abschlußbericht einer Arbeitsgruppe der Deutschen Forschungsgemeinschaft" ; *ZfBB - Zeitschrift für Bibliothekswesen und Bibliographie* ; Jg. 44 ; Nr. 1 ; S. 53-76.

ENDERLE W. (1997): "Auf dem Weg zur digitalen Bibliothek : Projekte in Deutschland" ; *Bibliothek* ; Jg. 21 ; Nr. 3 ; S. 359-362.

GABEL G. (1998): "DOI : Die ISBN des 21. Jahrhunderts?" ; *Bibliotheksdienst* ; Jg. 32 ; Nr. 6 ; S. 1069-1071.

Der Text ist zusätzlich als HTML-Datei verfügbar unter:

http://www.dbi-berlin.de/dbi_pub/bd_art/98_06_04.htm

- GEIBELMANN F., KECKS R.G. (2000): "Porträts im Internet" ; *BFB - Bibliotheksforum Bayern* ; Jg. 28 ; Nr. 2 ; S. 204-216.
- LESKIEN H. (2000): "Retrodigitalisierung : Eine Zwischenbilanz" ; *BFB - Bibliotheksforum Bayern* ; Jg. 28 ; Nr. 2 ; S. 132-153.
- MCMURDO G. (1997): "Web Graphics for the Lazy" ; *Journal of Information Science* ; Jg. 23 ; Nr. 2 ; S. 149-162.
- MEISSNER R. (1996): "Mischobst : Grafikformate für WWW-Dokumente" ; *c't - Magazin für Computer Technik* ; Nr. 7/96 ; S. 206 ff.
Der Text ist zusätzlich als HTML-Datei verfügbar unter:
<http://www.heise-online.de/ct/96/07/206/default.shtml>
- MEYENBURG S. (2000): "Der Aufbau Virtueller Fachbibliotheken in der Bundesrepublik Deutschland" ; *Bibliotheksdienst* ; Jg. 34 ; Nr. 7/8 ; S. 1229-1235.
Der Text ist zusätzlich als HTML-Datei verfügbar unter:
http://www.dbi-berlin.de/dbi_pub/bd_art/bd_2000/00_07_11.htm
- PETERS K. (1998): "Aus der Rechtskommission : Rechtsfragen der Bestanderhaltung durch Digitalisierung" ; *Bibliotheksdienst* ; Jg. 32 ; Nr. 11 ; S. 1949-1955.
Der Text ist zusätzlich als HTML-Datei verfügbar unter:
http://www.dbi-berlin.de/dbi_pub/bd_art/98_11_07.htm
- PUGLIA S. (2000): "Chapter VI : Technical Primer" ; *Handbook for Digital Projects : A Management Tool for Preservation and Access* ; SITTS M.K. [Hrsg.] ; Northeast Document Conservation Center ; Andover, Massachusetts.
Der Text ist zusätzlich als HTML-Datei verfügbar unter:
<http://www.nedcc.org/digital/VI.htm>
- REINITZER S., KROLLER F. (1996): "Langfristige Erhaltung digitaler Dokumente : Datenmigration oder Kulturverlust" ; *ABI-Technik* ; Jg. 16 ; Nr. 3 ; S. 278-282.
- SITTS M.K. [Hrsg.] (2000): *Handbook for Digital Projects : A Management Tool for Preservation and Access* ; Northeast Document Conservation Center ; Andover, Massachusetts.
Der Text ist zusätzlich als HTML-Datei verfügbar unter:
<http://www.nedcc.org/digital/TofC.htm>
- SMA (1999): "SMA Document-Hybridsystem für österreichische Nationalbibliothek - Scannen und Verfilmen in einem Arbeitsschritt" ; *ABI-Technik* ; Jg. 19 ; Nr. 1 ; S. 46-47.
- SMITH LEVINE M. (2000): "Chapter V : Overview of Legal Issues for Digitization" ; *Handbook for Digital Projects : A Management Tool for Preservation and Access* ; SITTS M.K. [Hrsg.] ; Northeast Document Conservation Center ; Andover, Massachusetts.
Der Text ist zusätzlich als HTML-Datei verfügbar unter:
<http://www.nedcc.org/digital/V.htm>
- TRENDS & MÄRKTE (1999): "Metasprache fürs Internet-Publishing : Flexibles Format" ; *Börsenblatt für den Deutschen Buchhandel* ; Nr. 49 / 22. Juni 1999 ; S. 10-12.
- VAN DER WERF T. (1998): *Metadata and Libraries : Background and Introduction* ; Text ist als RTF-Datei verfügbar unter:
<http://www.konbib.nl/persons/titia/publ/ticer.rtf>