

Durchführung von Digitalisierungsprojekten in Bibliotheken

Thomas Parschik

Der vorliegende Artikel entstand auf Grundlage einer Diplomarbeit,¹ in der ein Konzept für ein Digitalisierungsprojekt im Ibero-Amerikanischen Institut entwickelt wurde. Das Institut plant die Digitalisierung einiger Sammlungen lateinamerikanischer Volksliteratur und Grafik. Auch wenn die Entwicklung momentan noch im Fluss ist, sollen im nachfolgenden Artikel für Digitalisierungsprojekte wichtige Aspekte wie Vorgänge, Techniken und Standards² vorgestellt werden, allerdings unter Ausklammerung der umfassenden Themenkomplexe Metadaten und Präsentation der fertigen Ergebnisse im Internet.

1. Projektplanung

Die Projektplanung ist die Grundlage der Steuerung und Kontrolle eines Projekts. Unter Berücksichtigung der Aspekte Umfang, Zeit und Kosten und ggf. anderer Rahmenbedingungen beschreibt die Projektplanung, wie die Projektziele erreicht werden sollen. Zunächst werden in einem Lastenheft die Rahmenbedingungen geschildert, also die Anforderungen an den Leistungsumfang. Die Ziele werden definiert, nicht aber die Wege dorthin. Wichtige Endpunkte der Strukturplanung sind die Definierung der Projektstruktur, der Objektstruktur und der Arbeitsabläufe. Anschließend wird eine Projektorganisation festgelegt. In der Ressourcenplanung werden Personal und Hilfsmittel ausgewählt. Sodann wird ein Termin- bzw. Ablaufplan für die einzelnen Projektleistungen erstellt. Im Anschluss daran wird eine Risikobewertung vorgenommen. Risiken sind unter anderem eine unklare Definierung der Ziele und Anforderungen, ungeeignete Rahmenbedingungen, Probleme mit Mitarbeitern und technische Schwierigkeiten. Danach wird die Kostenplanung erstellt, d.h. die im Projektverlauf anfallenden Kosten abgeschätzt. Dann erfolgt die Leistungsbeschreibung in Gestalt eines Pflichtenheftes, das die Wege zur Realisierung der im Lastenheft definierten Anforderungen beschreibt.³

- 1 Thomas Parschik: Entwicklung eines Konzepts für ein Projekt zur Digitalisierung lateinamerikanischer Volksliteratur und Graphik im Ibero-Amerikanischen Institut Preußischer Kulturbesitz, Diplomarbeit an der Fachhochschule Potsdam, Fachbereich Informationswissenschaften, 2006
- 2 Digitalisierung von wissenschaftlichen Bibliotheken (2000, 2–4).
URL: <http://www.blk-bonn.de/papers/heft84.pdf> [Stand: 28.06.2006]
- 3 se: software-engineering; www.softwarekompetenz.de
URL: <http://www.software-kompetenz.de/?6156> [Stand: 28.06.2006]

1.1 Wirtschaftlichkeitsaspekte

Vor Projektbeginn muss eine Kostenkalkulation erstellt werden. Nur ein Teil der Kosten kann relativ genau im voraus berechnet werden (Personal, Geräte, Software, Speicherung), andere Kosten sind nur schwer abschätzbar (Gerätefehler, Strukturentwicklungen, kontinuierlicher Verwaltungsaufwand). Im Regelfall bildet das Personal den größten Kostenpunkt.⁴ Die Kosten sind vom Digitalisierungsumfang, der Scanauflösung und Farbtiefe, der Scantechnik, der Qualität der Vorlagen und der aktuellen Marktlage abhängig. Dazu kommen die Kosten für Bildbearbeitung, Indizierung der Images, Speicherung und Programmierung.⁵ Effektive Abläufe sind wichtig, da jede manuell durchgeführte Tätigkeit zusätzliche Kosten erzeugt. Auch treten Folgekosten für die in regelmäßigen Abständen erforderlichen Migrationen auf. U.U. ist eine Neudigitalisierung vom Mikrofilm preiswerter als eine Migrierung der alten Digitalisate.⁶ Da der Digitalisierungsaufwand sehr hoch ist, besteht in vielen Fällen die Notwendigkeit, Drittmittel einzuwerben.

Verfilmungs- und Digitalisierungsleistungen können intern durchgeführt oder outgesourct werden. Die Entscheidung hierüber wird u.a. von der Empfindlichkeit und Größe der Ausgangsmédien sowie vom Ausrüstungsbedarf beeinflusst. Wird die Verfilmung und Digitalisierung in der Bibliothek selbst vorgenommen, müssen die erforderlichen technischen Geräte gekauft oder gemietet werden, was evtl. teurer ist als Outsourcing.⁷ Für den Transfer und die Verarbeitung großer Datenmengen, wie sie bei der Digitalisierung anfallen, sind Standard-Bibliotheks-IT-Ausstattungen oft nicht ausreichend. Wird bei einer Verfilmung oder Digitalisierung in der Bibliothek zusätzliches Personal eingesetzt, erhöht sich kurzfristig der Projektoutput, die Effizienz aber sinkt, da die Mitarbeiter erst zeitaufwendig eingearbeitet werden müssen.⁸ Daher ist es oft sinnvoller, eine Fremdfirma zu beauftragen, die in diesem Bereich entsprechende Erfahrungen besitzt.

4 Bryant / Dobbie / Froud (2004, 46–47)

URL: http://www.bertelsmann-stiftung.de/cps/rde/xbcr/SID-0A000F0A-8CC32BC7/bst/Digitalisierung_kultureller_Ressourcen.pdf [Stand: 28.06.2006]

5 Maier / Exner (2000, 223)

6 Digitalisierung gefährdeter Bibliotheks- und Archivguts: digitale Beiträge zu archivischen Fragen Nr. 1; Abschlussbericht der Arbeitsgruppe „Digitalisierung“ des Unterausschusses Bestandserhaltung der Deutschen Forschungsgemeinschaft (7. Oktober 1996).

URL: http://www.landesarchiv-bw.de/sixcms/detail.php?template=hp_artikel&id=6511&id2=&sprache=de [Stand: 28.06.2006]

7 Bryant / Dobbie / Froud (2004, 37), s. Fußnote 4

8 Bryant / Dobbie / Froud (2004, 48), s. Fußnote 4

1.2 Arbeitsabläufe

Für die Auswahl von Beständen gibt es verschiedene Kriterien. Sie sollte die Bedürfnisse der Benutzer berücksichtigen⁹ und die Bedeutung der Bestände und dem Forschungsinteresse, das an der Nutzungsintensität abgelesen werden kann. Eventuell bietet sich die Möglichkeit, an bereits existierende digitale Angebote anderer Einrichtungen mit eigenen digitalen Beständen anzuknüpfen. Auch selten genutzte Bestände können ausgewählt werden, um sie einem breiten Publikum zugänglich zu machen. Oft werden besonders wertvolle oder unikale Bestände digitalisiert. Die Digitalisierung kann auch dem Schutz von im physischen Erhalt gefährdeten Beständen dienen.

Nach der Auswahl müssen die Größe und der Erhaltungszustand der Bestände festgestellt werden. Außerdem muss ermittelt werden, wie mit Besonderheiten wie beispielsweise Bindefehlern und fehlenden Seiten zu verfahren ist.¹⁰

Das zu digitalisierende Material wird zusammengestellt und einer Materialanalyse und -prüfung unterzogen. Die Ergebnisse entscheiden darüber, ob schwarz-weiß, in Graustufen oder in Farbe digitalisiert wird. Dann stellt man Testvorlagen für eine Probedigitalisierung zusammen. Entsprechend der Ergebnisse werden technische Parameter und Anforderungen an unterschiedliche Bildformate festgelegt. Wenn eine Fremdverfilmung und -digitalisierung vorgesehen ist, müssen Angebote von Digitalisierungsunternehmen eingeholt und miteinander verglichen werden. Das Ergebnis des Vergleichs ist die Auswahl eines Anbieters.

Die Form der Datenlieferung wird mit dem Anbieter vereinbart. Es werden Verzeichnisstrukturen und Dateinamen festgelegt, Konkordanzen zwischen Imagefolgen und festgelegten Strukturen erstellt und Metadaten für die bibliografische, inhaltliche und technische Beschreibung erstellt beziehungsweise Verfahren zu deren Übernahme aus vorhergehenden Systemen entwickelt. Ein Übergabeprotokoll muss erstellt und ein Verfahren zur Qualitätskontrolle entwickelt werden. Die manuelle Qualitätskontrolle ist sehr zeitaufwendig, da sie auch die Bildladezeit mit einschließt. Die Ergebnisse der Verfilmung und Digitalisierung müssen ständig kontrolliert werden, um Fehler und Mängel schnellstmöglich erkennen zu können. Je früher ein Fehler erkannt wird, um so niedriger sind die Kosten für seine Beseitigung.¹¹ Schließlich werden die Bilddateien ins eigene System importiert, gespeichert und archiviert.¹²

9 Bryant / Dobbie / Froud (2004, 24), s. Fußnote 4

10 Angaben von Herrn Rosenau, Geschäftsführer der MikroUnivers GmbH, gemacht im Interview am 26.06.2006 in Berlin

11 Bryant / Dobbie / Froud (2004, 48), s. Fußnote 4

12 Dörr (2000, 108–110)

Die Digitalisierung von Beständen ist nur sinnvoll, wenn die Digitalisate auch angemessen erschlossen werden. Es wird darüber entschieden, in welcher Struktur und in welchem Format die Digitalisate erstellt werden sollen. Falls sich bei vergleichbaren Projekten Standards bewährt haben, können diese übertragen werden. Weiterhin wichtig sind die personellen Ressourcen, die für die Erschließung eingesetzt werden können und sollen.

Das Projekt wird in einzelne Arbeitspakete eingeteilt, die aufeinander aufbauen oder parallel zueinander verlaufen.¹³ Alle Aufgaben müssen miteinander koordiniert werden, damit man flexibel reagieren kann, wenn ein Projektabschnitt früher als geplant abgeschlossen wird, oder wenn es zu unvorhergesehenen Verzögerungen kommt. Zeitliche und finanzielle Reserven sollten eingeplant werden, besonders dann, wenn Technologien eingesetzt werden, mit denen die Bibliothek noch nicht viel Erfahrung hat.¹⁴ Wird der Zeitrahmen für einen Arbeitsschritt oder für das gesamte Projekt zu eng gefasst, müssen bei der Zieldefinition Abstriche gemacht werden. Darum ist es wichtig, für das Projekt einen angemessenen Zeitrahmen festzulegen.¹⁵

2. Anforderungskatalog an Digitalisate

Vor der Probedigitalisierung wird ein Anforderungskatalog erstellt, in dem alle relevanten Informationen festgelegt werden. Dieser enthält folgende Angaben:

- Dateiformat (z.B. TIFF, JPEG, PDF)
- Auflösung
- Digitalisierung in Schwarz-Weiß, Graustufen oder Farbe
- beim TIFF-Format eine Indexierungsanleitung für den Header
- Speicherungs- und Lieferungsmodalitäten.

Soll zugleich verfilmt werden, müssen Filmformat (z.B. 35 mm) und Filmmaterial (Diazo oder Silberhalogenid) bestimmt, Bildlage und zu belassender Rand angegeben werden. Außerdem ist anzugeben, ob im Vollschrift oder im Halbschrift verfilmt werden soll. Gemäß DIN ISO 9001 wird ein Filmprotokoll erstellt.¹⁶ Die Entscheidung darüber, ob in Schwarz-Weiß, in Graustufen oder in Farbe und mit welcher Auflösung digitalisiert werden soll, hängt von den Anforderungen ab, die an die Digitalisate gestellt werden. Besonders hohe Anforderungen werden häufig an Digitalisate gestellt, die die Scanvorlagen für die Benutzung ersetzen sollen. Mit der Auflösung nehmen die Dateigröße und die Übertragungsdauer vom Scanner in den Computer zu. Dies zieht auch hohen Speicherplatzbedarf bei der

13 Exner (2000, 113–114)

14 Bryant / Dobbie / Froud (2004, 58–59), s. Fußnote 4

15 Bryant / Dobbie / Froud (2004, 46), s. Fußnote 4

16 Angaben von Herrn Rosenau (08.05.2006)

Archivierung nach sich. Hier müssen Aufwand und Nutzen sorgfältig abgewogen werden. Die Lesbarkeit eines Dokuments ist in der Regel mit einer Auflösung von 150–200 dpi in Graustufen gegeben. Bei normaler Schriftgröße kann mit einer Auflösung über 300 dpi in Graustufen keine signifikante Qualitätsverbesserung mehr erreicht werden. Weist die Vorlage eine geringe Schriftgröße auf, sollte mit 400 dpi gescannt werden, so dass die Zeichen im Digitalisat detailreicher dargestellt werden, was bes. beim Einsatz von OCR wichtig ist.¹⁷ Die Deutsche Forschungsgemeinschaft empfiehlt in von ihr unterstützten Digitalisierungsprojekten bei bitonalem Digitalisieren eine Auflösung von 600 dpi. Die Untergrenze liegt bei 300 dpi. Bei Graustufendigitalisaten empfiehlt die Deutsche Forschungsgemeinschaft (DFG) eine Auflösung von 300 dpi. Die Obergrenze liegt hier bei 600 dpi. Bei Farbdigitalisaten werden 600 dpi empfohlen. Die Digitalisate sollen mindestens Qualitätsindex 5 entsprechen.¹⁸

3. Gründe für parallele Digitalisierung und Mikroverfilmung

Durch Digitalisierung werden die zeitlichen und räumlichen Begrenzungen der Zugänglichmachung gedruckter Werke beseitigt.¹⁹ Höchstmögliche Wiedergabequalität, Verfügbarkeit, Zugänglichkeit und Wirtschaftlichkeit sollten gewährleistet sein. Die Digitalisierung erfüllt diese Kriterien noch nicht in zufriedenstellendem Maße. Die dauerhafte Erhaltung ist infolge des beständigen Wandels von Dateiformaten problematisch, der eine regelmäßige Migration in neue Formate erfordert. Sie kann durch Übertragung auf einen alterungsbeständigen Ersatzträger gewährleistet werden.²⁰ Die Mikroverfilmung ist neben der Erhaltung der Originalmedien die meistgenutzte Verfahrensweise zur dauerhaften Bestandssicherung.²¹ Mikrofilme sind keinem technischen Wandel unterworfen, unterliegen einer dichten nationalen und internationalen Normung, können kostengünstig hergestellt und dupliziert werden und erreichen eine hohe Wirtschaftlichkeit und Wiedergabequalität. Sie können erforderlichenfalls mit Mikrofilmscannern

17 Angaben von Herrn Dr. Dürr, Zentral- und Landesbibliothek Berlin, gemacht im Interview am 16.11.2006

18 Digitalisierungsrichtlinien der Deutschen Forschungs-Gemeinschaft, II.3.1.1: Digitalisierungsparameter URL: http://www.landesarchiv-bw.de/sixcms/detail.php?template=hp_artikel&id=6511&id2=&sprache=de [Stand: 28.06.2006]

19 Leskien, Hermann (o. J.): Die retrospektive Digitalisierung löst und impliziert Probleme. URL: <http://www.uni-muenster.de/Forum-Bestandserhaltung/konversion/digi-leskien.shtml> [Stand: 28.06.2006]

20 Digitalisierung gefährdeter Bibliotheks- und Archivguts, s. Fußnote 6

21 Hühn (2005, 23)

unter geringem Kostenaufwand (re-)digitalisiert werden.²² Allerdings wird die Benutzung von Mikrofilmen von vielen Benutzern als unangenehm empfunden. Manche Mikrofilmlesegeräte sind kompliziert zu handhaben. Mikrofilme können nur in der Bibliothek benutzt werden. Ihre Reproduktion ist mit Kosten- und Zeitaufwand verbunden.²³ Seit 2006 existieren Mikrofilmscanner, die Remote Access Service (RAS) möglich machen, d.h. sie ermöglichen den Zugriff auf Dokumente aus der Ferne. Der Benutzer wählt sich über ein analoges Modem oder über ISDN ein und identifiziert sich mit Benutzernamen und Passwort.²⁴

In der Fachliteratur wird empfohlen, das Original zu Sicherungszwecken zu verfilmen und auf der Grundlage des Mikrofilms ein Digitalisat zur Benutzung zu erstellen.²⁵ Einige Digitalisierungsunternehmen vertreten die Ansicht, dass eine bessere Qualität erreicht werden kann, wenn die Digitalisierung zuerst erfolgt und auf der Grundlage derselben der Mikrofilm erstellt wird.²⁶ Mit Hybrid-Verfahren können Verfilmung und Digitalisierung parallel erfolgen,²⁷ bis vor kurzen konnten Hybridgeräte nur Graustufenscans erstellen.²⁸

4. Mikroverfilmung

4.1 Anforderungen an Mikrofilme

Für kontrastreiche Vorlagen werden AHU-Mikrofilme auf Polyesterbasis empfohlen. Für die Verfilmung mit Graustufen werden häufig AHU-Halbtonefilme positiver Polarität benutzt, z.B. Kodak 2468 oder Kodak 3468. Während der Verfilmung müssen die Vorlagen vollständig und gleichmäßig ausgeleuchtet werden, um eine optimale Lesbarkeit zu gewährleisten. Diese bemisst sich am Qualitätsindex 8 auf Grundlage der DIN ISO 6199, Anhang C. Die erforderliche Auflösung wird an der Höhe des Kleinbuchstabens *e* folgendermaßen berechnet:

Geg.: *a* = Auflösung in Lp/mm (Linienpaare pro Millimeter);

h = Höhe des Kleinbuchstabens *e* in mm

Ges.: QI = Qualitätsindex

Berechnung des Qualitätsindex: $QI = a \times h$

22 Digitalisierung gefährdeten Bibliotheks- und Archivgut, s. Fußnote 6

23 Hühn (2005, 24)

24 Das ELKO: das Elektronik-Kompendium.

URL: <http://www.elektronik-kompendium.de/sites/net/0907081.htm>

[Stand: 28.06.2006]

25 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

26 Angaben von Herrn Brunow, Vertriebsmitarbeiter der Mik-Center GmbH, (28.04.2006) und von Herrn Rosenau (08.05.2006)

27 Hühn (2005, 31)

28 Angaben von Herrn Rosenau (08.05.2006)

Empfehlenswert sind 35-mm-Rollfilme, da sie für problematische Vorlagen bis zur Maximalgröße von 60 x 80 cm Qualitätsreserven besitzen. Handelsüblich sind Filmängen von 30,5 m und 65 m. Der Einsatz des letzteren ist effektiver, da eine größere Menge verfilmt werden kann, bevor der Film gewechselt werden muss. Ist der Mikrofilm als Grundlage für eine Digitalisierung vorgesehen, ist zu beachten, dass Digitalisate von Mikrofilmen negativer Polarität qualitativ besser sind als von Positivfilmen erstellte Digitalisate. Da jede Generierung einen geringen Qualitätsverlust mit sich bringt, sollte ein Mikrofilm niedriger Generationsstufe als Digitalisierungsgrundlage verwendet werden. Der sogenannte Preservation Master, (Film 1. Generation) dient der Sicherung und darf nicht als Grundlage der Digitalisierung verwendet werden. Hierfür sollte ein Duplikat des Preservation Master in negativer Polarität auf Silberhalogenid- oder Diazoform verwendet werden. Für farbige Digitalisate ist ein hochauflösender und farbbeständiger Farbmikrofilm zu empfehlen, der nach dem Farbbleichverfahren auf Polyesterbasis hergestellt wurde. Die Erstellung von Farbmikrofilmen erfolgt meist mit einem proprietären System. Damit wird der digitalisierte Bildausschnitt auf 24 x 36 mm reduziert. Im Vollschrift beträgt der Bildausschnitt einer Farbmikrofilmaufnahme 32 x 45 mm. Vor der Erstellung von Halbschrittaufnahmen und kleineren Bildfeldern ist zu prüfen, ob diese vom System übertragbar sind, da Rollfilme nur in einer Richtung eingeführt werden können. Auch unzerschnittene Mikrofilme können eingeführt werden, was aber mit der Gefahr der Beschädigung des Films verbunden ist, wenn der Mikrofilmsscanner nicht über Aufspulvorrichtungen verfügt. Die Filmvorlage wird digitalisiert und komprimiert auf einer Foto-CD gespeichert.²⁹

4.2 Der Mikroverfilmungsprozess vom Original

Beim Verfilmungsprozess sollten die Vorlagen gleich gestaltet sein, da jede Umjustierung den Arbeitsaufwand erhöht. Man sollte einen einheitlichen Verkleinerungsfaktor wählen. Im Header des Bildformates werden die Auflösung und die Gesamtzahl der Pixel erfasst. Die Vorlagen müssen einheitlich und lesegerichtet auf dem Aufnahmetisch positioniert werden. Die gewählte Positionierung sollte während des Verfilmungsprozesses nicht verändert werden. Die Aufnahme im Vollschrift in Bildlage 2 A gemäß DIN ISO 6199 wird empfohlen. Die Seiten können mittig getrennt werden.³⁰ Dies wird Halbschrittverfilmung genannt. Da mehr Bildstege vorhanden sind, passen weniger Aufnahmen auf einen Film. Auf einen 30,5 m langen Film passen je nach Vorlage ca. 620–630 Halbschrittaufnahmen. Bei Vollschriftverfilmung passen auf einen 30,5 m langen Film ca. 1.200 Aufnahmen.³¹ Bei manuellem mittigem Trennen ist die Genauigkeit größer als bei maschinell

29 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

30 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

31 Angaben von Herrn Brunow (28.04.2006)

vorgenommenem mittigem Trennen. Letzteres setzt eine einheitliche Positionierung der Vorlage voraus. Wenn der Film nur für Sicherungszwecke vorgesehen ist, kann man in Bildlage 2 A gemäß DIN 19057 verfilmen. Dabei liegt die Aufnahme hochkant. So passen je nach Vorlage ca. 2.400 Aufnahmen auf einen 35-mm-Film bzw. ca. 8.000 Aufnahmen auf einen 16-mm-Film. Der 16-mm-Film ist platzsparend und dem 35-mm-Film qualitativ gleichwertig. Er wird in Deutschland jedoch nicht als Standard anerkannt. Die DFG fordert in von ihr finanzierten Projekten die Verwendung von 35-mm-Filmen.³² Bei paralleler Digitalisierung und Mikroverfilmung können die gescannten Bilder mit einer Langzeitarchivierungssoftware auf den Mikrofilm übertragen werden, bspw. mit dem ArchiveWriter OP 500 von Zeutschel. Er hat eine Verarbeitungsgeschwindigkeit von bis zu 1.200 Images pro Stunde.³³ Am Anfang des Mikrofilms muss ein ausreichender Vorspann belassen werden, um einen Eigentumsnachweis anbringen und eine problemlose Einlegbarkeit in das Mikrofilmlesegerät gewährleisten zu können. Auf dem Mikrofilm und der Mikrofilmschachtel sind äußerlich die Filmnummer und alle inhaltlich relevanten Angaben zu vermerken.³⁴ Auf Mikrofilmen kann man Bildmarken (Blips) setzen. Auch Gruppenblips für mehrere Aufnahmen können vergeben werden. Bildmarken erhöhen die Zugriffsgenauigkeit erreicht. Der Aufwand für das Setzen von Bildmarken ist verhältnismäßig gering.³⁵

5. Der Digitalisierungsprozess

Es gibt keine allgemeingültigen Digitalisierungsstandards. Fortwährende Veränderungen von Datenbankarchitekturen, Dateiformaten und Übertragungsgeschwindigkeiten machen eine beständige Anpassung der bestehenden Standards erforderlich. Darum sollten digitale Dokumentenbestände so angelegt werden, dass man sie einheitlich im Stapelbetrieb ändern kann.³⁶

Vor Beginn der Digitalisierung wird mit ausgewählten Digitalisierungsvorlagen ein Anforderungskatalog an die Testdigitalisate erstellt, der alle für die Digitalisierung erforderlichen Informationen, die technischen Angaben, beispielsweise die Auflösung und das Dateiformat, die formale Materialbeschreibung und eine Indexierungsanleitung sowie Angaben über die Form der Langzeitarchivierung enthält. Eine Indexierungsanleitung wird erstellt, die von der Dokumentenstruktur der Vorlage abhängig ist. Sie definiert die zu übernehmenden Strukturelemente und ermöglicht so die spätere Aufbereitung der Metadaten und der Erschließungs-

32 Angaben von Herrn Rosenau (08.05.2006)

33 Zeutschel. URL: <http://www.zeutschel.de/produkte/pdf/op500.pdf> [Stand: 28.06.2006]

34 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

35 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

36 Bryant / Dobbie / Froud (2004, 53), s. Fußnote 4

struktur. Jede Vorlage muss mehrfach digitalisiert werden, um aufwendige Nacharbeiten zu vermeiden.³⁷

Die Qualität eines Digitalisats wird am Qualitätsindex (QI) gemessen. Der gebräuchliche QI liegt zwischen 3 und 9. Eine mittlere Qualität ist mit einem QI mit 5, eine hohe Qualität mit einem QI um 8 gegeben.³⁸ Die Qualität eines Digitalisats hängt von der Tonalität des Scanners und von der Bildauflösung ab. Die Tonalität gibt die Anzahl der Bits an, mit denen die Helligkeitsstufe oder Farbzusammensetzung eines Bildpunktes beschrieben wird. Je höher die Anzahl der Bits ist, um so größer ist auch die Farbgenauigkeit des Digitalisats. Die Bildauflösung wird in Dots per Inch (dpi) gemessen. Damit wird die Anzahl der Pixel pro Zoll (2,54 cm) angegeben. Man unterscheidet optische und interpolierte Bildauflösung. Die optische Auflösung ist von der Anzahl der Fotosensoren eines CCD-Zeilensensors und dem Zeilenvorschub abhängig. Bei der interpolierten Auflösung werden zusätzliche Bildpunkte erzeugt. Aus den angrenzenden Bildpunkten werden Mittelwerte gebildet. Mit der so ermittelten Farbe werden die zusätzlichen Bildpunkte ausgefüllt.³⁹ Ob schwarz-weiß, in Graustufen oder in Farbe digitalisiert wird, ist unter anderem vom zu digitalisierenden Medium abhängig. Wenn die Vorlage Bilder enthält oder der Kontrast zwischen Schrift und Hintergrund gering ist, sollte in Graustufen digitalisiert werden. In Farbe wird digitalisiert, wenn in der Vorlage Farben einen Informationswert besitzen, oder wenn die Lesbarkeit wesentlich verbessert wird.⁴⁰ Von kontraststeigernden AHU-Mikrofilmen wird mit 16 Graustufen digitalisiert. Bei Halbtonverfilmungen wird mit 256 Graustufen digitalisiert. Hierfür ist ein Speicherplatz von 8 Bit erforderlich.⁴¹ Die erforderliche Auflösung richtet sich nach der Höhe des Kleinbuchstabens *e* in mm. Sie wird folgendermaßen berechnet:

Geg.: a = Auflösung in dpi ; h = Höhe des Kleinbuchstabens *e* in mm

Ges.: QI = Qualitätsindex

Berechnung des Qualitätsindex bei bitonalem Digitalisieren:

$$QI = (a \times 0,039h) / 3$$

Berechnung des Qualitätsindex für Digitalisierung mit Graustufen:

$$QI = (a \times 0,039h) / 2^{42}$$

37 Hühn (2005, 34–36)

38 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

39 Rollnik (1999, 19–20)

40 Angaben von Herrn Rosenau (26.06.2006)

41 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

42 Rautenberg / Titel (2003, 53)

URL: <http://www.buchwiss.uni-erlangen.de/AllesBuch/Susanne%20Starnes/Starnes.pdf> [Stand: 28.06.2006]

Bei bitonalem Digitalisieren sind Auflösungen zwischen 350 und 400 dpi, beim Digitalisieren mit Graustufen 250 bis 300 dpi gebräuchlich. Wenn ein Mikrofilm als Speichermedium dient, muss nicht die höchstmögliche Wiedergabequalität des Digitalisats erreicht werden. Bei einer Digitalisierung vom Original hat das Digitalisat die Funktion des Speichermediums und muss besonders hohe Anforderungen erfüllen, da es ggf. das Original ersetzen muss. Das Digitalisat muss mindestens den QI 8 erreichen.⁴³ Auch die Ausbelichtung hat einen Einfluss auf die Qualität des Digitalisats. Bei Vorlagen, die Halbtöne enthalten, erfolgt eine weiche Ausbelichtung, eine sogenannte flache Gradation. Bei kontrastreichen Vorlagen wird eine harte Ausbelichtung, eine sogenannte steile Gradation, vorgenommen.⁴⁴ Für die Ausbelichtung kann bspw. der ArchiveLaser des Fraunhofer Instituts für physikalische Messtechnik verwendet werden. Für die Belichtung einer Vorlage mit 15.000 Zeilen benötigt er etwa 40 Sekunden.⁴⁵ Zur Schärfemessung werden Testmieren verwendet. Das sind Schablonen, die der DIN 19051 entsprechen und probeweise gescannt werden. Auf einer Testmiere sind Kreise in Gruppen zu 4 x 4 Kreisen angeordnet. In jedem Kreis verlaufen zwei Striche parallel zueinander. Die Schärfe wird in dpi gemessen.⁴⁶ Die Messung der optischen Dichte erfolgt mit einem Densitometer, einer aus verschiedenfarbigen Feldern bestehenden lichtdurchlässigen Schablone. Auch das Densitometer entspricht der DIN 19051. Die Mikrofilmdichte wird mit einem Graukeil gemessen. Das ist ein Filmstück, das verschiedene Dichtestufen enthält.⁴⁷

5.1 Scanner

Die Qualität von Digitalisaten hängt von der Qualität des verwendeten Scanners ab. Bei der Auswahl eines Scanners ist darauf zu achten, dass dieser unter Berücksichtigung der Qualität der Scanvorlagen alle gestellten Anforderungen erfüllt. Ein Scanner ist ein Lesegerät, das Vorlagen in eine maschinenlesbare Form überträgt. In dieser Form bestehen die Bilder aus einzelnen Pixeln, deren Dichte über die Schärfe des Bildes entscheidet. Der Scanner ist als Peripheriegerät an den Computer angeschlossen. Er sollte über eine SCSI-Schnittstelle verfügen.⁴⁸ SCSI

43 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

44 Angaben von Herrn Rosenau (08.05.2006)

45 Fraunhofer Institut für physikalische Messtechnik.
URL: <http://www.ipm.fraunhofer.de/fhg/ipm/extra/bigimg/laserbelichtung/archivlaser/archivlasergeraetgr.jsp> [Stand: 28.06.2006]

46 Angaben von Herrn Rosenau (08.05.2006)

47 Verband der Mikrofilm-Fachbetriebe e.V. (1987, T. 2, 106)

48 Bericht der Arbeitsgruppe Technik zur Vorbereitung des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“ im Förderbereich „Verteilte Digitale Forschungsbibliothek“.

(Small Computer Systems Interface) ist ein Peripheriebus für den Anschluss von Scannern, Massenspeicherlaufwerken und Peripheriegeräten. SCSI-Schnittstellen verfügen über eine hohe Übertragungsrate. Sie sind weit verbreitet, flexibel und leicht zu konfigurieren.⁴⁹ Für die Auswahl des Scanners ist die Übertragungsgeschwindigkeit von Bedeutung.⁵⁰ Man unterscheidet Scanner mit CCD-Technik und Scanner mit Multiplierteknik.⁵¹ Ein CCD (Charge Coupled Device) ist ein Halbleiterdetektor, der aus einer Matrix lichtempfindlicher Zellen (Pixel) besteht. Der CCD-Chip einer Digitalkamera besitzt zwischen 300.000 und 1.000.000 Pixel.⁵² Die meisten mit CCD-Technik ausgerüsteten Scanner verfügen über einen beweglichen CCD-Zeilensensor, auch CCD-Array genannt, der aus 2.000 bis 4.000 linear angeordneten Fotosensoren besteht. Beim Scanvorgang wird die Scanvorlage von einer integrierten Lichtquelle beleuchtet. Das Licht wird von der Vorlage reflektiert und über ein Spiegel- und Linsensystem zu den Fotosensoren geleitet, die das so reflektierte Bild in einzelne Lichtpunkte zerlegen und die Intensität des reflektierten Lichts erfassen. Gemäß dieser Lichtstärke nehmen die Fotosensoren eine entsprechend hohe oder niedrige Spannung an. Diese Spannungen werden sodann von einem Analog/Digitalwandler in ganze Zahlen übersetzt und an einen Computer geschickt, wo diese Werte im RAM-Speicher erfasst werden.⁵³ Beim Scanvorgang mittels CCD-Zeilensensor soll die Vorlage vor IR-Licht, UV-Licht und Wärme geschützt werden. Die für den Scanvorgang erforderliche Lichtmenge wird möglichst gering gehalten und die Dauer der Lichteinwirkung auf eine Stelle der Vorlage verkürzt. Im Unterschied zum CCD-Scanner ist die Vorlage bspw. beim Aufsichtsscanner wesentlich länger der Beleuchtung ausgesetzt.⁵⁴ Um Farbdigitalisate herzustellen, müssen Farbfilter in den Scanner eingesetzt werden, die nur Lichtstrahlen eines bestimmten Wellenlängenbereichs durchlassen. Sogenannte Dreidurchgangsscanner (Three-Pass-Scanner) arbeiten mit CCD-Technik. Sie tasten die Digitalisierungsvorlage in drei Scanvorgängen mit einer Scanzeile mit weißem Licht ab und erfassen bei einem Durchgang die roten, bei einem Durchgang die grünen und bei einem Durchgang die blauen Bildteile. Anschließend

URL: http://www.user.gwdg.de/~sub/ebene_2/vdf/entwurf3.htm [Stand: 28.06.2006]

49 Transtec: the European IT Factory; maßgeschneiderte IT-Systeme.

URL: <http://www.transtec.de/D/D/IT-Kompodium/ITKnowHow/Storagebuses/SCSIinterfaces.html> [Stand: 28.06.2006]

50 Bericht der Arbeitsgruppe Technik zur Vorbereitung des Programms „Retrospektive Digitalisierung von Bibliotheksbeständen“, s. Fußnote 48

51 Rollnik (1999, 16–17)

52 Lexikona : Enzyklopaedie & Wissen. URL: http://charge-coupled_device.lexikona.de/art/Charge-coupled_Device.html [Stand: 28.06.2006]

53 Rollnik (1999, 16–17)

54 Angaben von Herrn Rosenau (26.06.2006)

werden die Farbauszüge deckungsgleich übereinander gelegt. So entsteht ein farbiges Abbild der Vorlage. Schneller, aber auch teurer, sind Eindurchgangsscanner (Single-Pass-Scanner), die alle drei Farbanteile in einem Scanvorgang erfassen.⁵⁵ Ob ein Scanner für ein Digitalisierungsprojekt geeignet ist, richtet sich u.a. nach der Verarbeitungsgeschwindigkeit. Diese wird in DIN-A4-Blättern pro min. gemessen, wobei man von einer Auflösung von 200 dpi ausgeht.⁵⁶

Einzugsscanner sind in der Lage, schnell und selbstständig große Mengen zu digitalisieren. Die Digitalisierungsvorlage wird in das Gerät eingezogen und an einem mit CCD-Zeilen ausgerüsteten Fenster vorbeigeführt. Dies schließt den Einsatz für das Digitalisieren gebundener Vorlagen aus. Auch empfindliche, insbesondere brüchige, Papiere sollten nicht mit einem Einzugsscanner digitalisiert werden.

Bei Flachbettscannern wird das zu digitalisierende Dokument auf eine Glasplatte aufgelegt. Sodann wird der Scanner mit einem Deckel über der Scannvorlage geschlossen, und eine Abtastzeile leuchtet die Vorlage ab. Flachbettscanner erreichen eine Auflösung bis zu 2.400 dpi. Sie sind für DIN-A4- und DIN-A3-Formate geeignet. Mit Flachbettscannern ist die Digitalisierung einzelner Blätter, Akten, Bücher und Zeitschriften möglich. Bei gebundenen Vorlagen wird allerdings der Einband belastet. Außerdem ist die Belastung der Vorlage durch UV-Licht und Wärme sehr groß. Deshalb ist der Einsatz von Flachbettscannern für schutzwürdige Vorlagen auszuschließen.

Aufsichtsscanner, auch Buchscanner genannt, wurden für gebundene Vorlagen wie Bücher, Zeitschriftenbände und Akten sowie für Landkarten entwickelt.⁵⁷ Sie werden u.a. von den Firmen Cruse⁵⁸, Image Ware⁵⁹ und Zeutschel⁶⁰ hergestellt. Für großformatige Vorlagen werden Scanner eingesetzt, die mit CCD-Kameras ausgerüstet sind. In der Gehäuserückwand der Kamera sind matrixförmig CCD-Zellen angeordnet. CCD-Kameras können eine Auflösung bis zu 1.200 dpi erreichen. Um bei minimaler Belastung der Einbände ein Digitalisat von maximaler Qualität zu erhalten, wird die Falzwölbung durch den Einsatz einer Buchwippe ausgeglichen.⁶¹ Erforderlichenfalls kann durch zusätzliche Hilfsmittel das Buch nur in einem 90°-Winkel geöffnet und gescannt werden.

55 Rollnik (1999, 17–19)

56 Rollnik (1999, 25)

57 Billy ([2003], 30–31)

58 Cruse : Digital Equipment. URL: <http://www.crusedigital.com/scanners.html> [Stand: 28.06.2006]

59 ImageWare. URL: <http://www.imageware.de/de/systeme/> [Stand: 28.06.2006]

60 Zeutschel. URL: <http://www.zeutschel.de/produkte.html> [Stand: 28.06.2006]

61 Billy ([2003], 33)

5.2 Speicherung

Für die Datenübertragung werden DAT-Medien (Digital Audio Tapes) und CD-R (Compact Disc Recordable) empfohlen, die hardwareunabhängig lesbar sind. Die CD-R sind vorzuziehen, da sich bei den DAT die Spannung der Bänder verändern kann. Außerdem kann man auf Digitalisate der CD-R direkt zugreifen, ohne diese zuvor auf der Festplatte speichern zu müssen. DVD-R (Digital Versatile Disc Recordable) bieten gegenüber der CD-R zwar eine deutliche höhere Speicherkapazität, sind aber aufgrund der höheren Beschreibdichte empfindlicher. Alle diese Träger garantieren nicht die Langzeitarchivierung.

Ein Digitalisat ist erst dann ausreichend gesichert, wenn es auf einen Primärdatenträger und ein Arbeitsduplikat übertragen wurde, deren Lesbarkeit und inhaltliche Konsistenz überprüft wurden. Ein Dekompressionstest kann durchgeführt werden.⁶² Unter Dekompression versteht man die Wiederherstellung von mittels Kompressionsverfahren reduzierten Daten.⁶³ Die Digitalisate müssen in gängigen Formaten vorliegen.⁶⁴ Beim Digitalisieren werden üblicherweise ein Masterimage und ein Arbeitsimage erstellt. Das Masterimage muss eine möglichst große Wiedergabequalität gewährleisten. Das Arbeitsimage dient der Bereitstellung an die Benutzer.⁶⁵

Die Formate GIF und JFIF sind verlustbehaftet und damit ungeeignet.⁶⁶ JPEG entspricht der ISO/IEC IS 10918. Sein Gebrauch wird von der International Telecommunication Union (ITU) in der Recommendation T.81 empfohlen. JPEG kann hohe Farbtiefen speichern, ist aber verlustbehaftet.⁶⁷ Einige Digitalisierungsfirmen sind der Ansicht, dass die Qualität von JPEG für die Langzeitarchivierung ausreicht.⁶⁸ Die DFG ist der Ansicht, dass JPEG sich lediglich als Bereitstellungsformat eignet.⁶⁹ PNG verfügt über eine hohe Farbtiefe und schnelle Komprimierungsalgorithmen.⁷⁰ PNG-Dateien sind oftmals größer als JPEG-Dateien der gleichen Wiedergabequalität. Das PNG-Format ist nicht sehr verbreitet, und nur wenige Programme kön-

62 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

63 Documanager.de : das Dokumenten Management Portal.
URL: http://www.documanager.de/ressourcen/glossar_468_dekompression.html
[Stand: 28.06.2006]

64 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

65 Angaben von Herrn Rosenau (26.06.2006)

66 Rohde-Enslin (2004, 16, 14)

67 Rohde-Enslin (2004, 14)

68 Angaben von Herrn Rosenau (08.05.2006)

69 Digitalisierungsrichtlinien der Deutschen Forschungs-Gemeinschaft, II.3.3.3 (a) JPEG

70 Hühn (2005, 33)

nen PNG-Dateien erstellen.⁷¹ Die Möglichkeiten der Abspeicherung zusätzlicher Informationen in der Imagedatei sind sehr begrenzt.⁷² Einige Digitalisierungsunternehmen vertreten die Auffassung, dass PNG die Anforderungen an Digitalisate zur Langzeitarchivierung erfüllt.⁷³ Die DFG akzeptiert die Verwendung von PNG in von ihr geförderten Projekten nur als Bereitstellungsformat,⁷⁴ nicht für die Langzeitarchivierung.⁷⁵ Mit TIFF ist die Wiedergabe hoher Farbtiefen möglich. Es können Dateien bis zu einer maximalen Größe von 2^{32} Bytes erstellt werden. In anderen Formaten vorliegende Daten können innerhalb des TIFF-Formats gespeichert werden. TIFF ist frei benutzbar. Auf der Grundlage von TIFF können neue Programme erstellt werden.⁷⁶ TIFF ist strukturiert aufgebaut. Es besteht aus einem Header und einer beliebigen Anzahl von Datenblöcken variabler Länge, die mittels Zeiger adressiert werden. In die Image File Directory können zusätzliche Daten wie beispielsweise Größe, Auflösung, Farbtiefe oder Metainformationen eingegeben werden. TIFF bietet über 90 Kategorien (Tags) an. Die Verwendung einiger dieser Kategorien wird von der DFG empfohlen.⁷⁷ TIFF ist plattformunabhängig und kann mit unterschiedlichen Betriebssystemen und Programmen weiterverarbeitet werden. TIFF-Dateien können unkomprimiert oder komprimiert bezogen werden. Schwarz-Weiß-Digitalisate können mit TIFF G 4 verlustfrei komprimiert werden. Der Bezug komprimierter Dateien erlaubt die Einsparung von Speicherplatz. Zuvor muss sichergestellt sein, dass die Programme des Empfängers mit komprimierten TIFF-Dateien arbeiten können.⁷⁸ Es wird empfohlen, TIFF-Dateien unkomprimiert zu speichern.⁷⁹ Gemäß den Richtlinien der DFG ist das TIFF-Format für die Langzeitarchivierung zu empfehlen.⁸⁰

Das Format PDF wurde von der Firma Adobe entwickelt. Es ist weit verbreitet und plattformunabhängig. In PDF-Dokumente bleiben Layout, Schriftarten, Bilder und Verlinkungen unveränderlich erhalten. Es können auch Hypertext-Links, Stichwortsuche und Multimedia-Elemente integriert werden. Allerdings ist die Rückumwandlung in Word-Dokumente problematisch.⁸¹ Bei PDF-Dateien können

71 Rohde-Enslin (2004, 18)

72 Hühn (2005, 33)

73 Angaben von Herrn Rosenau (08.05.2006)

74 Digitalisierungsrichtlinien der Deutschen Forschungs-Gemeinschaft, II.3.3.3 (b) PNG

75 Digitalisierungsrichtlinien der Deutschen Forschungs-Gemeinschaft, II.3.1.2 (b) PNG

76 Rohde-Enslin (2004, 12–13)

77 Hühn (2005, 32)

78 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

79 Rohde-Enslin (2004, 13)

80 Digitalisierungsrichtlinien der Deutschen Forschungs-Gemeinschaft, II.3.1.2 (a) TIFF

81 PDF- Portable Document Format. URL: <http://www.hrz.uni-wuppertal.de/infos/hrz-info/hrz-info-200307/node64.html> [Stand: 15.11.2006]

Digitalisate anstelle von Einzelseiten beispielsweise zu Heft- oder Bandstrukturen gebündelt werden. Die Daten lassen sich gut komprimieren, allerdings sind solche PDF-Dateien dennoch größer als Einzelimages und benötigen eine entsprechend längere Bildladezeit. PDF-Dateien bieten auch die Möglichkeit, Bild- und Textdaten zu kombinieren. In einer unsichtbaren Hintergrundebene wird hinter dem angezeigten Bild ein OCR-Ergebnis abgelegt, so dass das angezeigte Faksimile zugleich als Volltext recherchierbar ist. Das PDF-Format bietet auch differenzierte Möglichkeiten des Schutzes vor urheberrechtswidriger Nutzung, beispielsweise durch die Vergabe von Passwörtern oder von Wasserzeichen. Das PDF-Format ist inzwischen in der Variante PDF/A einer der Standards für digitale Archivierung.⁸²

6. Erschließung der Digitalisate

Die Digitalisate sollten formal und inhaltlich erschlossen werden. Die formale Erschließung erfolgt i.d.R. nach den RAK-WB, v.a. wenn die Digitalisate im Bibliothekskatalog recherchierbar sein sollen. Die inhaltliche Erschließung trägt zur Schaffung eines Mehrwertes bei. Schon eine einfache inhaltliche Erschließung durch Vergabe von Suchkriterien erleichtert den Überblick über die Bestände zu einem bestimmten Thema. Eine sehr tiefgehende Form der Erschließung ist der Einsatz automatischer Texterkennung, der es möglich macht, mittels Volltextsuche große Bestandsmengen innerhalb von Sekunden auch nach nicht indexierten Begriffen zu durchsuchen. Der Mehrwert wächst mit dem Grad der Erschließung.⁸³

Man unterscheidet Digitalisate mit kodierten Informationen (Coded Information, CI) und Digitalisate mit nicht kodierten Informationen (Non Coded Information, NCI). Bei kodierten Digitalisaten ist jedes Schriftzeichen mit einem Code abgespeichert, der dem PC eine Interpretation ermöglicht. Speicherung im Volltext ermöglicht eine Volltextsuche. Für Text- und Metadaten haben sich XML-basierte Datenformate durchgesetzt, wobei von der DFG die Ausprägung METS (Metadata Encoding and Transmission Standard) empfohlen wird.⁸⁴ Nicht kodierte Bild-Informationen können zum Beispiel in TIFF, JPG oder PDF vorliegen. Solche Digitalisate können vom PC nicht interpretiert werden.⁸⁵

82 PDF als Archivierungsstandard : pdf als Grundlage für die Archivierung von Papier und elektronischen Dokumenten. URL: <http://www.adobe.com/de/products/acrobat/pdfs/pdfarchiving.pdf> [Stand: 16.11.2006]

83 Leskien, s. Fußnote 19

84 Digitalisierungsrichtlinien der Deutschen Forschungsgemeinschaft. URL: http://66.249.93.104/search?q=cache:ap6zeeYbCHQJ:www.dfg.de/forschungsfoerderung/formulare/download/12_151.rtf+bildlage+2A+2B&hl=de&gl=de&ct=clnk&cd=1 [Stand: 19.06.2006]

85 Rollnik (1999, 10–11)

Für die Anzeige im Internet ist es in vielen Fällen sinnvoll, digitalisierte Bilddaten in drei Fassungen bereit zu stellen: in einem kleinen Vorschaubild, das dem Benutzer einen ersten Eindruck vermittelt, und dessen Dateigröße zwischen 5 und 20 KB liegt, in einer Vollansicht unter Ausnutzung der gesamten Bildschirmgröße, deren Größe zwischen 30 und 120 KB liegt, und in einer Detailansicht, die zwischen 100 und 250 KB liegt und auch Details deutlich erkennbar macht. Die beiden letzteren können auch identisch sein, das heißt, dass die Vollansicht so weit vergrößert werden kann, dass man nur noch einen Teil des Bildes, diesen aber dafür sehr deutlich, sieht.

Für die Präsentation müssen über die bibliografischen Daten hinaus verschiedene Metadaten vorgehalten werden: die Signatur des Mediums, formale Kriterien wie die Größe des Originals, Material, Schäden und Restaurationen des Originals,⁸⁶ aber auch technische Daten wie Bildgröße, Auflösung, Format und Bittiefe.⁸⁷ Digitalisierte Bestände müssen auch nach bestimmten Gliederungskriterien erschlossen sein. Thematisch zusammengehörende Dokumente werden in Sammlungen gegliedert. Alle zu einem Dokument gehörenden Digitalisate benötigen eine gemeinsame Markierung. Metadaten können auf verschiedene Weise erzeugt werden, worauf im folgenden näher eingegangen werden soll.

6.1 Manuelle Indexierung

Die qualitativ höchstwertige Erschließung ist die manuell vorgenommene Indexierung. Sie setzt ein inhaltliches Verständnis des digitalisierten Dokuments voraus. Dabei muss eine Indexierungssprache angewendet werden, in der die relevanten Inhalte des zu erschließenden Dokuments wiedergegeben werden.⁸⁸ Das „DFG-Förderprogramm retrospektive Digitalisierung“ vertritt die Auffassung, dass die inhaltliche Erschließung der Digitalisate finanziell nachrangig zu behandeln sei. In einigen US-amerikanischen Digitalisierungsprojekten wurde ein Zeitraum von 15 Minuten für die manuelle Erschließung eines Dokumentes veranschlagt.⁸⁹

6.2 Automatische Schrifterkennung

Eine preiswerte Erschließungsmöglichkeit bietet der Einsatz von Automatischer Schrifterkennung (Optical Character Recognition, OCR). Es handelt sich hierbei um ein automatisches Verfahren, das mittels Mustererkennung bildlich vorliegende alphanumerische Zeichen in codierte Daten umwandelt. Zunächst wird

86 Uhde (2000, 240)

87 Digitalisierung gefährdeter Bibliotheks- und Archivguts, s. Fußnote 6

88 Nohr (2003, 19)

89 Leskien, s. Fußnote 19

die Vorlage gescannt und ein Digitalisat erstellt. Anschließend führt das Schrift-erkennungsverfahren eine Dokumentenstrukturanalyse durch. Dabei werden das Layout, der Satzspiegel und die Zeilenposition sowie Grafiken und Tabellen analysiert. Dann erfolgt die Bestimmung der Richtung der Zeichenabfolge und eine Begradigung der Zeichenkonturen, um eine bessere Erkennung zu ermöglichen. Die Schriftzeichen werden voneinander separiert, indem ein Feldmuster auf den Text gelegt wird. Unter Einsatz von Algorithmen wird das Digitalisat in einen editierbaren ASCII-Text umgewandelt. Es folgt die Abschlusskontrolle, die vom Gerät oder manuell durchgeführt werden kann.⁹⁰ Einige Systeme arbeiten mit unscharfer Logik (fuzzy logic). Sie setzen bei nicht eindeutig identifizierbaren Schriftzeichen das Zeichen ein, dessen Verwendung am wahrscheinlichsten ist. Andere Verfahren arbeiten mit Mixed Mode. Sie fügen nicht erkannte Zeichen als Images in den Text ein. OCR erkennt und berücksichtigt Textspalten, Textblöcke und Grafiken. Der Einsatz von OCR wird als wirtschaftlich angesehen, wenn die Erkennungsgenauigkeit bei 99,95% liegt (max. 5 Fehler pro 1.000 Zeichen).⁹¹ Eine so hohe Erkennungsrate ist praktisch nur bei manueller Erfassung oder manueller Nachkorrektur erreichbar. In der Realität erreichen automatische Texterkennungsprogramme eine Erkennungsgenauigkeit von maximal 98–99%.⁹² Die gewöhnliche vollautomatische Texterkennung ermöglicht lediglich die Volltextsuche.⁹³ Einige Texterkennungsprogramme sind in der Lage, Dokumente unter Einsatz verschiedener Methoden inhaltlich zu erschließen. Die Erkennungsrate ist von der Qualität der OCR-Software und der Beschaffenheit der Vorlage sowie vom angewandten Digitalisierungsverfahren abhängig. Ist die Vorlage stark verschmutzt oder zerknittert, ist die Gefahr der Fehlinterpretation von Zeichen sehr hoch. Auch die Schriftart der Vorlage beeinflusst die Erkennungsgenauigkeit. Man unterscheidet Schriften mit fester Zeichenbreite und Proportionalschrift mit variabler Zeichenbreite. Die Schrifterkennungssoftware orientiert sich an den Abständen zwischen den einzelnen Buchstaben. Bei Schreibmaschinenschrift werden die Schriftzeichen in horizontaler Richtung in einheitlichen Abständen gesetzt. Sie ist für Schrifterkennungsprogramme leichter zu interpretieren als beispielsweise am PC erstellte Dokumente in Proportionalschrift, da hier eine exakte Zeichen-segmentierung nicht immer möglich ist. Da die OCR-Software ein Feldmuster auf den Text legt, kann ein uneinheitlicher Schriftzeichenabstand zur Folge haben, dass mehrere Zeichen einem Feld zugeordnet werden. Auch der Schrifttyp ist mit entscheidend für die Wirtschaftlichkeit des Einsatzes von OCR. Sogenannte Designerschrift erschwert den Einsatz von OCR. Einige Schriftzeichen sehen sich sehr

90 Rollnik (1999, 31–32)

91 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

92 Angaben von Herrn Rosenau (26.06.2006)

93 Digitalisierung gefährdeten Bibliotheks- und Archivguts, s. Fußnote 6

ähnlich und werden deshalb oft falsch interpretiert, beispielsweise die römische Zahl I, der Großbuchstabe I und der Kleinbuchstabe i.⁹⁴

Ferner haben die Scanauflösung, die Farbtiefe, der Schwellenwert und nicht zuletzt die Qualität der eingesetzten Texterkennungssoftware Einfluss auf die Erkennungsgenauigkeit. Diese kann durch Trainieren des Texterkennungsprogramms erhöht werden.⁹⁵ Um die Interpretierbarkeit zu verbessern, sollten kleine Schriften mindestens mit 400 dpi gescannt werden. Die beim Einsatz von automatischer Schrifterkennung auftretenden Fehler werden in falsch interpretierte Zeichen und nicht zuordenbare Zeichen unterschieden. Falsch interpretierte Zeichen werden nicht automatisch angezeigt. Sie müssen nach dem automatischen Schrifterkennungsprozess manuell gesucht werden oder bleiben bestehen. Nicht zuordenbare Zeichen werden zur Nachbereitung automatisch zur Kontrolle angezeigt. Eine große Erkennungsgenauigkeit ist mit ABBYY FineReader 8.0 gewährleistet, andere verbreitete Programme sind Omnipage Professional 15 und Readiris Pro 11. Ein besonderes Problem stellen Frakturschriften dar. Die Firma ABBYY hat die Software Fine Reader XIX entwickelt, die auf Druckschriften zwischen 1880 und 1938 spezialisiert ist und Frakturschrift, Schwabacher, Textur- und romanische Schrift erkennt.⁹⁶ Eine weitere Texterkennungssoftware für Frakturschrift ist das von der Firma B.I.T. entwickelte Programm BIT-Fraktur.⁹⁷

6.3 Statistische Indexierungsverfahren

Statistische Indexierungsverfahren gehen davon aus, dass die Häufigkeit des Auftretens eines Wortes in einem Dokument Aufschluss über die Signifikanz desselben gibt (Zipfsches Gesetz). Ein Begriff, der in einem zu erschließenden Dokument häufig, im Gesamtbestand der Dokumente selten vorkommt, ist von inhaltlicher Relevanz. Ein Begriff, der in einem Dokument häufig und auch im Gesamtbestand der Dokumente häufig auftritt, ist nicht von inhaltlicher Relevanz. Da nicht alle Worte sinntragend und daher als Indexterme geeignet sind, muss eine Indexterminologie erstellt werden, in die nur geeignete Terme aufgenommen werden. Indexterme haben verschiedene Wertigkeiten. Folglich muss eine Gewichtung vorgenommen werden.⁹⁸

94 Rollnik (1999, 32–34)

95 Fricke / Maier (2000, 201–202)

96 CHIP ONLINE DE.

URL: http://www.chip.de/news/c1_news_14490699.html?tid1=28882&tid2=0
[Stand: 28.06.2006]

97 B.I.T. Bureau Ingénieur Tomasi erhält Fraktur-Auftrag von ZLB
In: *Bibliotheksdiens*t, 2006, Nr. 10, S. 1187–1188

98 Nohr (2003, 33–34)

6.4 Linguistische Indexierungsverfahren

Linguistische Verfahren können Flexionsformen erkennen (Maskulinum, Femininum, Neutrum, Singular, Plural). Eine Rechtschreibprüfung ermöglicht das Auffinden eines Begriffs auch bei falscher oder abweichender Schreibung bei der Recherche oder im digitalen Dokument. Man unterscheidet folgende Fehlerarten:

- die Auslassung eines Zeichens, das in den Begriff gehört
- die Einfügung eines Zeichens, das nicht in den Begriff gehört
- die Substitution, also die Ersetzung eines Zeichens, das in den Begriff gehört, durch ein anderes Zeichen, das nicht in den Begriff gehört
- die Vertauschung von Zeichen in ihrer Reihenfolge.

Bei der syntaktischen Analyse werden Begriffe auf ihre Grundformen reduziert. Für jeden Begriff muss festgelegt werden, wie weit er reduziert werden soll. Wird der Stamm zu weit belassen, können bei der Recherche viele irrelevante Treffer angezeigt werden (Overstemming). Wird der Stamm zu knapp bemessen, werden evtl. relevante Dokumente nicht gefunden (Understemming). Zudem dient die syntaktische Analyse der Erkennung von Homographen, indem sie Mehrwortgruppen erkennt. Vor der Anwendung eines syntaktischen Verfahrens muss ein Wörterbuch erarbeitet werden.⁹⁹

Informationslinguistische Indexsysteme werden in regelbasierte und wörterbuchbasierte Verfahren eingeteilt. Bei ersteren muss ein Regelwerk erstellt werden. Alle in den digitalisierten Dokumenten auftretenden Wörter werden gemäß diesem Regelwerk analysiert und bearbeitet. Dieses Verfahren ist fehlerbehaftet. Das regelbasierte Verfahren ist mit einem verhältnismäßig geringen Aufwand verbunden, da hierbei nicht jeder relevante Begriff einzeln bearbeitet wird. Wörterbuchbasierte Verfahren sind sehr zuverlässig. Sprachliche Unregelmäßigkeiten werden ausgeglichen. Die Erarbeitung von Wörterbüchern ist extrem arbeits-, zeit- und kostenaufwendig.¹⁰⁰ Außerdem orientiert sich ein Wörterbuch zumeist an aktuellen sprachlichen und orthografischen Regelungen und erkennt alte Formen nicht.¹⁰¹

Morphologische Verfahren basieren auf den Gesetzmäßigkeiten von Wortstrukturen und der Bildung von Wortklassen. Man unterscheidet Flexionsmorphologie und Derivationsmorphologie. Bei der Flexionsmorphologie werden Wörter umgewandelt, um grammatikalische Kontraste innerhalb von Satzkonstruktionen auszudrücken. Die Derivationsmorphologie befasst sich mit Grundprinzipien der Konstruktion neuer Wörter.¹⁰²

99 Nohr (2003, 47–52)

100 Nohr (2003, 55–56)

101 Fricke / Maier (2000, 221)

102 Nohr (2003, 54–55)

6.5 Mustererkennungsverfahren

Mustererkennungsverfahren (Pattern-Matching-Verfahren) können Muster nicht nur in Texten, sondern auch in anderen Dokumentformen erkennen. Ein lernfähiges System kann entwickelt werden, das im Laufe der Zeit durch manuell vorgenommenes Indexieren lernt, Dokumente in definierte Klassen einzuordnen. Es wird ein Abgleich zwischen sprachlichen Mustern in den digitalisierten Dokumenten mit Einträgen in einer Wissensbasis vorgenommen. Pattern-Matching-Verfahren besitzen ein hohes Weiterentwicklungspotential, sind derzeit aber noch nicht weit verbreitet.¹⁰³

7. Weitere Faktoren bei der Projektplanung

7.1 Schutz der Digitalisate

Sofern eine Bereitstellung auf Open-Access-Basis aus rechtlichen Gründen nicht möglich ist oder von der anbietenden Bibliothek nicht gewünscht wird, müssen digitale Bestände vor unberechtigter Reproduktion und Nutzung geschützt werden. Eine Manipulation des Bibliotheksservers ist i.d.R. nicht möglich. Der Schutz von im Internet angebotenen Digitalisaten vor unberechtigter Vervielfältigung und Verwendung ist problematisch. Mit dem Programm Java Script kann die Funktion „Speichern unter“ blockiert werden. Dies verhindert jedoch nicht, dass das Digitalisat über die Cachefunktion kopiert werden kann. Die Cachefunktion speichert jede von einem PC aufgerufene Internetseite ab, um sie bei einem wiederholten Seitenaufruf schneller anzeigen zu können. Diese Funktion kann nicht blockiert werden. Ein Digitalisat kann mit einem digitalen Wasserzeichen versehen werden, das einen Teil des Bildes verdeckt.¹⁰⁴ Hierfür kann bspw. das kostenlose Programm „Bildschutz Pro“¹⁰⁵ verwendet werden.¹⁰⁶ Eine Alternative sind die Schutzmöglichkeiten, die das PDF-Format bietet. Eine weitere Möglichkeit, Digitalisate vor unerlaubter Vervielfältigung zu schützen, besteht darin, diese in so geringer Qualität anzubieten, dass sie für eine Nachnutzung nicht brauchbar sind. Ein absoluter Schutz kann nur durch die ausschließliche Bereitstellung im Intranet gewährleistet werden.¹⁰⁷

103 Nohr (2003, 55–56)

104 Angaben von Herrn Lozze, Mitarbeiter des Ibero-Amerikanischen Instituts Preußischer Kulturbesitz, gemacht am 23.06.2006

105 URL: <http://www.webprofi-tools.de/bspro.htm>

106 KLAB development.

URL.: <http://www.webprofi-tools.de/bspro.htm> [Stand: 28.06.2006]

107 Angaben von Herrn Lozze (23.06.2006)

7.2 Urheberrecht

Das Urheberrecht schützt das Recht eines Urhebers an seinem Werk bzw. des Verlegers, dem Nutzungsrechte übertragen wurden. Demgegenüber hat die Allgemeinheit ein Interesse an der Zugänglichkeit von Informationen sowohl für private und berufliche Zwecke als auch für Bildung und Wissenschaft. In der Bundesrepublik Deutschland sind die Rechte der Rechteinhaber durch Art. 14 Abs. 2 GG dann begrenzt, wenn ein überwiegendes Bedürfnis der Allgemeinheit vorliegt. Man unterscheidet zwischen einem besonders privilegierten Zweck und dem überwiegenden Allgemeininteresse. Somit können Werke ggf. auch ohne Zustimmung und Vergütung des Urhebers genutzt werden.¹⁰⁸

§ 53 Abs. 2 Nr. 2 UrhG erlaubt die Vervielfältigung eines Werkes „zur Aufnahme in ein eigenes Archiv, wenn und soweit die Vervielfältigung zu diesem Zweck geboten ist und als Vorlage für die Vervielfältigung ein eigenes Werkstück benutzt wird.“ Die Vervielfältigung kann selbst oder von einem Dritten im Auftrag durchgeführt werden. Laut § 53 Abs. 5 UrhG sind „Datenbankwerke, deren Elemente einzeln mit Hilfe elektronischer Mittel zugänglich sind“, hiervon ausgeschlossen. Urheberrechtlich geschützte Werke oder Teile von Werken dürfen demgemäß nur dann in Datenbanken verwendet werden, wenn „der wissenschaftliche Gebrauch sowie der Gebrauch im Unterricht nicht zu gewerblichen Zwecken“ gewährleistet ist. Demzufolge muss das Nutzungsrecht erworben werden. Gelingt dies nicht, können die Digitalisate nur im Intranet bereitgestellt werden.¹⁰⁹ Werden ausländische Publikationen digitalisiert, gelten ggf. die gesetzlichen Bestimmungen des jeweiligen Landes.

Die Recherche nach Rechteinhabern ist sehr zeitaufwendig, mit ein Grund dafür, dass v.a. urheberrechtfreie Werke digitalisiert werden. Je mehr Rechteinhaber vorhanden sind, um so geringer ist die Chance, die Zustimmung aller Rechteinhaber für die Digitalisierung zu bekommen. Je älter ein Werk ist, desto schwerer ist es, den Rechteinhaber zu kontaktieren, v.a. nach dem Tod des ursprünglichen Rechteinhabers ist es mitunter nicht leicht, zu ermitteln, an wen die Rechte übergegangen sind.¹¹⁰ Nach deutschem Recht erlischt der Rechtsschutz ein Werkes 70 Jahre nach dem Tod aller Urheber.

Die Suche nach dem Rechteinhaber sollte folgendermaßen verlaufen: ein Arbeitsblatt wird erstellt, auf dem der Name des Rechteinhabers, das Anfangsdatum der Recherche und der Name des Bearbeiters sowie die betreffenden Werke

108 Beger (2002, 10–15)

109 Das deutsche Urheberrechtsgesetz – UrhG: Stand: September 2003; Abschnitt 6; Schranken des Urheberrechts.

URL: <http://transpatent.com/gesetze/urhg11.html#53> [Stand: 28.06.2006]

110 Bryant / Dobbie / Froud (2004, 40), s. Fußnote 4

aufgeführt werden. Anschließend erfolgt die Suche nach dem Rechteinhaber unter Einsatz von Print- und Onlinequellen, Branchenkontakten, Zeitungsmeldungen und Nachrufen. Stellt sich heraus, dass das Urheberrecht auf einen anderen Inhaber übergegangen ist, so wird dieser auf dem Formular angegeben. Ist der Rechteinhaber bereits verstorben, ist das Todesdatum zu notieren. Sodann folgt die Suche nach Angehörigen. Die Richtigkeit von ermittelten E-Mail-Adressen und Telefonnummern muss überprüft werden. Dann wird ein sogenanntes Copyright-Paket zusammengestellt und an den Rechteinhaber geschickt. Es enthält ein persönliches Anschreiben, ein Standardfreigabeformular und eine Liste der betreffenden Werke. Im Anschreiben bittet man den Rechteinhaber um Antwort innerhalb einer bestimmten Frist. Das Absendedatum wird auf dem Arbeitsblatt vermerkt. Geht seitens des Rechteinhabers innerhalb der gegebenen Frist keine Antwort ein, kann man nach Fristablauf telefonisch, per Fax, E-Mail oder Brief nachfragen. Datum und Art der Kontaktaufnahme sind ebenso wie der Name des Bearbeiters auf dem Arbeitsblatt zu vermerken. Bei telefonischen Kontakten müssen zudem der Name des Gesprächspartners und der Gesprächsinhalt vermerkt werden. Wird auf einem Anrufbeantworter eine Nachricht hinterlassen, muss auch diese inhaltlich auf dem Arbeitsblatt wiedergegeben werden. Bleibt wiederum eine Antwort aus, kann man nach Verstreichen einer angemessenen Zeit den Rechteinhaber nochmals kontaktieren. Eine dritte und vierte Nachfrage können in angemessenen zeitlichen Abständen folgen. Erfolgt keinerlei Interaktion seitens des Rechteinhabers, muss entschieden werden, ob man das Verhalten des Rechteinhabers als Freigabeverweigerung werten soll. Anderenfalls muss das Urheberrecht neu geprüft werden. Verweigert der Rechteinhaber die Freigabe des Vervielfältigungsrechts, dürfen die betreffenden Dokumente nicht ins Internet gestellt werden.¹¹¹ Sowohl die Suche nach den Rechteinhabern als auch jeglicher Kontakt mit denselben muss dokumentiert werden, das heißt, dass sämtliche Briefe, Notizen von Telefongesprächen und natürlich die Rechtfreigabe in einem Aktenordner gesammelt werden. Auch eine Rechteverweigerung muss aufbewahrt werden. Copyright-Management-Services können mit der Ermittlung der Copyrightverhältnisse beauftragt werden. Die kanadische Firma Access Copyright¹¹² ist auf die Suche von Copyrightinhabern spezialisiert.

Bei der Digitalisierung von Nachlässen sind daneben mitunter Persönlichkeitsrechte zu beachten.

111 Bryant / Dobbie / Froud (2004, 91), s. Fußnote 4

112 URL: <http://www.accesscopyright.ca/>

Literatur

Beger, Gabriele: Urheberrecht und elektronische Bibliotheksangebote : ein Interessenkonflikt. Berlin, 2002

Billy, Jonas Bakoubayi: Vorschlag für ein Pilotprojekt: Restaurierung, Digitalisierung und Mikroverfilmung von deutschsprachigen Unterlagen im Nationalarchiv der Republik Togo. Potsdam, [2003]

Dörr, Marianne: Planung und Durchführung von Digitalisierungsprojekten. In: Weber, Hartmut / Maier, Gerald [Hrsg.]: Digitale Archive und Bibliotheken : neue Nutzungsmöglichkeiten und Nutzungsqualitäten. Stuttgart. 2000, S. 103–112

Exner, Peter: Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut. In: Weber, Hartmut / Maier, Gerald [Hrsg.]: Digitale Archive und Bibliotheken : neue Nutzungsmöglichkeiten und Nutzungsqualitäten. Stuttgart. 2000, S. 113–127

Fricke, Thomas / Maier, Gerald: Automatische Texterkennung bei digitalisiertem Archiv- und Bibliotheksgut. In: Weber, Hartmut / Maier, Gerald [Hrsg.]: Digitale Archive und Bibliotheken : neue Nutzungsmöglichkeiten und Nutzungsqualitäten. Stuttgart. 2000, S. 201–221

Hühn, Chris: Digitalisierung und Nutzbarmachung von historischen Vorlesungsmitschriften : Diplomarbeit zur Erlangung des Grades einer Diplombibliothekarin (FH) eingereicht am Fachbereich Informationswissenschaften. Potsdam, 2005

Maier, Gerald / Exner, Peter: Wirtschaftlichkeitsüberlegungen für die Digitalisierung von Archiv- und Bibliotheksgut. In: Weber, Hartmut / Maier, Gerald [Hrsg.]: Digitale Archive und Bibliotheken : neue Nutzungsmöglichkeiten und Nutzungsqualitäten. Stuttgart. 2000, S. 223–229

Nohr, Holger: Grundlagen der automatischen Indexierung : ein Lehrbuch. Berlin, 2003

Rohde-Enslin, Stefan: Nicht von Dauer : kleiner Ratgeber für die Bewahrung digitaler Daten in Museen. Berlin, 2004

Rollnik, Ulrike: Retrospektive Digitalisierung von konventionellen Zeitungsausschnittbeständen : ein Überblick mit praktischen Beispielen des im Zeitungsausschnittarchiv der Hochschule für Film und Fernsehen „Konrad Wolf“ durchgeführten DEFA Scanprojektes ; Diplomarbeit zur Erlangung des Grades einer Diplomdokumentarin (FH), Fachbereich Archiv-Bibliothek-Dokumentation der Fachhochschule Potsdam. Potsdam, 1999

Uhde, Karsten: Kontextbezogene Online-Präsentationen von Archivgut. In: Weber, Hartmut / Maier, Gerald [Hrsg.]: Digitale Archive und Bibliotheken : neue Nutzungsmöglichkeiten und Nutzungsqualitäten. Stuttgart. 2000, S. 233–245

Verband der Mikrofilm-Fachbetriebe e. V.: Ausbildungsleitfaden Reprograf für die Aus- und Weiterbildung von Fachpersonal : Teil 2. Stuttgart, 1987