

95. Deutscher Bibliothekartag Dresden

24. März 2006

Themenkreis 3:

Erhaltung und Präsentation der kulturellen Überlieferung

9.00 Uhr

Titel des Referats:

**Bestandserhaltung mit Digitalisaten
in Der Deutschen Bibliothek - aktuelle
Überlegungen und Planungen**

Referent:

Dr. Nicolas Flores-Herr
Koordinator Digitalisierung
IT-Abteilung
Die Deutsche Bibliothek
Adickesallee 1
60322 Frankfurt am Main
Tel.: 069/1525-1705
email:flores@dbf.ddb.de

Motivation

Die Deutsche Bibliothek hat die Aufgabe, das wissenschaftliche und kulturelle Erbe Deutschlands zu bewahren und der Öffentlichkeit dauerhaft zugänglich zu machen. In zunehmenden Maße sollen außerdem Bestände retrospektiv digitalisiert werden, die aufgrund der Stofflichkeit des Trägermaterials mit fortschreitendem Alter zerfallen.

In digitaler Form sind sie für Nutzer "bis in Ewigkeit" verwendbar – natürlich nur unter bestimmten Voraussetzungen, während die gefährdeten Originale vor weiterer Benutzung verschont werden. Ein weiterer Nutzen ist die komfortable und effektive Recherche von Informationen zu Publikationen, die in digitaler Form vorliegen.

Eine der größten Herausforderungen bei Digitalisierungsvorhaben sind die entstehenden Kosten: sowohl in der Phase der Überführung des Originals in ein Digitalisat als auch bei deren Erschließung und Präsentation.

Daher sollte bereits in der Planungsphase der Mehrwert der Digitalisate im Hinblick auf folgende Aspekte bewertet werden.

- **Bewahrung des kulturellen Erbes**
 - Langzeitarchivierung Digitalisate
- **Mehrwert**
 - Forschung & Wissenschaft
 - Bildung
 - Kultur
- **Komfortable Nutzung**

Innerhalb dieses Kontextes sollte die Nutzbarkeit von hochqualitativer und hochgradig verfügbarer Information das Hauptmotiv und Stein des Anstoßes für das Digitalisierungsprojekt sein.

Um Digitalisate für Bestandserhaltung verwertbar zu machen, müssen diese einerseits in sicheren Speichersystemen vorgehalten werden als auch

langzeitarchiviert werden. Allein die Langzeitarchivierung gewährleistet die Lesbarkeit der Dateien ‚bis in Ewigkeit‘ - über die Bewahrung des Datenstroms hinaus - durch Migration der Dateiformate und Emulation aktuell bestehender Computersysteme.

Geschäftsgang

Für Digitalisate ist seitens der Die Deutsche Bibliothek folgender Geschäftsgang geplant.

- 1. Erzeugung**
- 2. Import**
 - a. Eintrag in Katalogsystem (Pica)
 - b. Vergabe von persistenten Identifikatoren (URN)
 - c. Erzeugung von abgeleiteten Formen (.gif, .jpeg, ...)
 - d. LZA System (kopal/DIAS)**
- 3. Erschließung**
 - a. Inhaltserfassung
 - b. Strukturerrfassung
- 4. Export**
 - a. Präsentation Lesesaal/Internet, Kunde
 - b. Export aus LZA System (kopal/DIAS)**
 - c. Schnittstellen Daten/Metadaten (Portale, Datendienste)

In den Schritten 1.-3. werden die Digitalisate mit Metadaten angereichert, die einerseits der komfortablen Nutzung dienen als auch deren Langzeitarchivierbarkeit sichern.

Automatisierung des Geschäftsgangs

Da der größte Teil der Digitalisierungskosten durch Personalaufwand verursacht wird, möchte Die Deutsche Bibliothek möglichst viele Schritte des Geschäftsgangs automatisieren. Die aktuelle Planung fasst folgende Bereiche des Digitalisierungs-Geschäftsgangs für eine Automatisierung ins Auge.

- Import
- Inhaltserfassung
- Strukturerrfassung
- Präsentation

Des Weiteren sollen schon bei der Erzeugung der Digitalisate mit Hilfe neuer Hochdurchsatz-Scantechnologie Kosten eingespart werden. Es bestehen Überlegungen den vollautomatischen Buchscanner APT 2400 der Firma Kirtas Technologies (Victor, NY; USA) im Rahmen eines Testprogramms zu evaluieren.

Messbasiertes Qualitätsmanagement

Im Folgenden wird erörtert, wie Information der Vorlage durch Digitalisierung gesichert werden soll. Durch neue Ansätze im Bereich Qualitätsmanagement sollen objektive Kriterien erstellt werden, welche zum Ziel haben bestehende – wahrnehmungsbasierte – Scanrichtlinien um messbasierte Vorschriften zu erweitern.

Die häufigsten Fehlerquellen beim Digitalisieren von Bildern können mit folgenden Scanparametern assoziiert werden.

- 1.) Abtastrate <> Scanauflösung
- 2.) Farbe <> Farbtiefe
- 3.) Optische Abbildung <> Geometrische Verzerrung

Diese Fehlerquellen können durch Messung von Ortsfrequenzgang, Dynamikumfang, Tonwertumfang und Druckbildgeometrie der Vorlage quantifiziert werden.

Die so gewonnenen Werte liefern beim Erstellen von messbasierten Qualitätskriterien ein objektives Maß, welcher Teil der Information der Vorlage bei gewählten Scanparametern digitalisiert wird und welcher nicht.

Die Scanparameter sollten dann derart gewählt werden, dass die gewünschte Information der Vorlage mit maximaler Qualität bei möglichst niedrigem Speicherverbrauch digitalisiert wird.

1.) Abtastrate - Scanauflösung

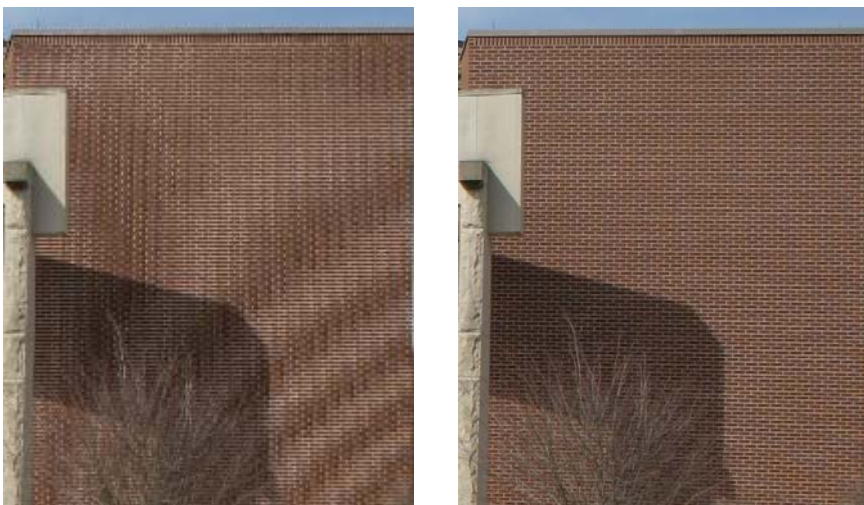
Mit welcher Abtastrate (f_{abast}) die Vorlage mindestens digitalisiert werden soll, wird durch das Abtasttheorem von Harry Nyquist und Claude Elwood Shannon spezifiziert.

$$f_{abast} > 2 \cdot f_{max}$$

Nyquist-Shannon-Abtasttheorem

Das Nyquist-Shannon-Abtasttheorem besagt, dass der Betrag der Abtastrate (f_{abast}) größer als das Doppelte der maximalen Ortsfrequenz (f_{max}) der Vorlage sein muss. Wird beim Digitalisieren unterhalb der doppelten Ortsfrequenz abgetastet ($f_{abast} < 2 \cdot f_{max}$), entstehen Aliasing-Artefakte, die sich vor allem durch Moiré-Muster bei Farb- oder Graustufenabbildungen oder durch treppenartige Strukturen bei Schwarzweissbildern bemerkbar machen.

Beispiel:



Fotos: Colin M.L. Burnett

Moiré-Effekt: Das Bild auf der linken Seite weist die für den Moiré-Effekt typischen Helligkeitsmodulationen auf, wie sie durch zu niedrige Abtastung beim Fotografieren des Mauerwerks hervorgerufen werden können. Das Bild auf der rechten Seite zeigt das gleiche Motiv mit ausreichender Abtastung digitalisiert.

In der Praxis würde die Messung der maximalen Ortsfrequenz jedes einzelnen Druckwerks zuviel Zeit in Anspruch nehmen, um praktikabel zu sein. Da die maximale Ortsfrequenz der gängigen Druckverfahren im Wesentlichen durch die minimale Rasterweite gegeben ist, muss nicht für jeden Vorlagentyp eine Messung durchgeführt werden.

Druckvorlage	Rasterweite (f_{\max})	Minimale Scanauflösung ($f_{\text{abta\text{st}}}$)
Zeitung	≤ 100 lpi	220 ppi
Zeitschrift	120 – 150 lpi	330 ppi
Buch	≤ 200 lpi	440 ppi

Scanauflösung: Rasterweiten gängiger Druckverfahren und die sich daraus ergebende Digitalisierungsauflösung¹

Diese Tabelle liefert ein objektives Maß, wie die Information, welche im Druckbild steckt, durch Wahl einer geeigneten Scanauflösung digitalisiert werden kann. Hierbei wird natürlich nicht berücksichtigt, wie detailreich die Strukturen anderer Bestandteile des Buchs (z.B. Papier, Einband) erfasst werden. Hierfür müsste in aufwendigen Verfahren die maximale Ortsfrequenz dieser Bestandteile gemessen werden.

2.) Farbe - Farbtiefe

Die Farbtiefe von Rasterdruckerzeugnissen muss ebenfalls nicht umständlich gemessen werden. Die Anzahl der Tonwerte bzw. Helligkeitsstufen eines Offsetdrucks ist im Regelfall nicht größer als 5-6 Bit pro Farbe. Durch den konsequenten Einsatz von Farbmanagement sowie der Verwendung von rauscharmen Sensoren mit entsprechendem Tonwertumfang und Dynamikbereich kann die Farbe der Vorlage problemlos in digitale Bildformate mit einer Farbtiefe von 8 Bit pro Farbkanal überführt werden.

3.) Optische Abbildung <> geometrische Verzerrung

Die Verzerrung der optischen Abbildung ist bei professionellen Flachbettscannern vernachlässigbar. Wird mit Hilfe von optischen Systemen (z.B. Digitalkameras) digitalisiert, muss die Kissenverzerrung des optischen Systems gemessen und herausgerechnet werden. Hierfür muss hauptsächlich die Verzeichnung der verwendeten Kameraobjektive (kissenförmig oder tonnenförmig) gemessen werden und daraus eine „Entzerrungs-Matrix“ errechnet werden. Mit Hilfe dieser Matrix kann geometrische Verzerrung aus dem digitalen Bild herausgerechnet werden.

Zusammenfassung

Die aktuellen Überlegungen von Die Deutsche Bibliothek zur Schonung angegriffener Bestände durch Digitalisierung basieren auf

1) Kostensenkung durch

¹ Die angegebenen Scanauflösungen beziehen auf das reale Auflösungsvermögen der gesamten Digitalisierungsapparatur (z.B. Scanner oder Digitalkamera) welches zunächst durch Messung der optischen Übertragungsfunktion (MTF; Modulation Transfer Function) bestimmt werden muss. Die MTF gibt weiterhin Aufschluss darüber, welche Qualität die Digitalisierungsapparatur bei der gegebenen Auflösung hat. In der Praxis (z.B. CD) wird für die Abtastung meist mit einem Faktor von 2,2 abgetastet.

- a. Hochdurchsatz-Scantechnologie
 - b. Automatisierung des Geschäftsgangs
- 2) Messbasierte Qualitätssicherung der Scanparameter
- a. Auflösung (ppi)
 - b. Farbtiefe (Bit)
 - c. Geometrische Verzerrung
- 3) Sicherung durch Langzeitarchivierung

Die Kostensenkung durch Hochdurchsatz-Scantechnologie soll sowohl durch den Einsatz von vollautomatischen Scanrobotern realisiert werden als auch durch die weitestgehende Automatisierung des Imports, der Inhalts- bzw. Struktur Erfassung und Präsentation der Digitalisate erfolgen. Um den Qualitätsstandard der Digitalisate objektiv bewerten zu können, wurden messbasierte Qualitätsrichtlinien erarbeitet. Die dauerhafte Nutzbarkeit der Bestände soll – bei gleichzeitiger Schonung und Bewahrung des Originals – durch Systeme zur Langzeitarchivierung und Sicherung von Digitalisaten (kopal/DIAS) gewährleistet werden.