

## Digitalisierung

### WORKSHOP DER MASSENDIGITALISIERUNGSPROJEKTE DER DEUTSCHEN FORSCHUNGSGEMEINSCHAFT AN DER HERZOG AUGUST BIBLIOTHEK WOLFENBÜTTEL

Vor zwei Jahren hat die DFG einen Förderschwerpunkt gebildet, der sich die Digitalisierung der im VD16 und VD17 nachgewiesenen Drucke zum Ziel gesetzt hat. Mittlerweile liegen erste belastbare Ergebnisse aus dieser Förderlinie vor. Dies bot Anlass und Gelegenheit, zu einem gemeinsamen Workshop nach Wolfenbüttel einzuladen, um Erfahrungen der Massendigitalisierungsprojekte auszutauschen und auszuwerten. Dabei sollten die praktischen Probleme im Vordergrund stehen, insbesondere Fragen zu Arbeitsabläufen, Kosten, Digitalisierungsverfahren, konservatorischen Problemen und Erschließungsmaßnahmen wie Katalogisierung, Strukturdatenerschließung und Tiefenerschließung. Die jeweiligen Themen wurden durch Initiativreferate skizziert und im Anschluss diskutiert. Der nachstehende Bericht fasst die Ergebnisse der eintägigen Veranstaltung vom 17. März 2009 zusammen.

Einleitend gab Thomas Stäcker (HAB Wolfenbüttel) einen Überblick über die in den Projekten geleisteten Arbeiten. Die Zusammenstellung basierte auf Statistikblättern, die vor der Tagung von den Teilnehmern erbeten worden waren und in der Regel den Stand vom 31. Januar 2009 widerspiegeln.

Erfragt wurden Angaben zur Imagedigitalisierung und Erschließung. Die Angaben zur Imagedigitalisierung (vgl. Abb. 2 zur Imagedigitalisierung) sind auf eine Seite heruntergerechnete Mittelwerte. Sie stellen daher keine exakte Kalkulation dar, wohl aber einen Orientierungswert. Das zeigt sich schon daran, dass die Daten, bei aller Interpretationsbedürftigkeit und Abhängigkeit von den jeweiligen Rahmenbedingungen, insgesamt eine gewisse Einheitlichkeit aufweisen. Interessant ist, dass die Kosten für Dienstleister und Inhouse-Lösungen dicht beieinander liegen. Ein erheblicher Kostenfaktor ist in vielen Fällen die konservatorisch bedingte einseitige Digitalisierung, die bestandsabhängig Mischkalkulationen erzwingt. Die Spalte »Aufnahmen« weist insofern nur das theoretische

Potential aus, das zu erreichen wäre, wenn Aufnahmetechnik und konservatorischer Zustand des Objektes eine doppelseitige Digitalisierung, d. h. die Erzeugung von zwei Bilddateien/Images mit nur einem Klick, erlaubten. Der Aufwand für die konservatorische Prüfung und die bibliografische bzw. Strukturdatenerschließung ließ sich aus den Statistikblättern nicht eindeutig ermitteln. Das liegt vor allem daran, dass Personal mit mehreren Aufgaben im Digitalisierungsprozess betraut wurde und Projektmitarbeiter neben Aufgaben zur Katalogisierung von Sekundärausgaben, auch Tätigkeiten bei der Strukturdatenerschließung und dem Projektmanagement übernahmen. Immerhin ist zu erkennen, dass Katalogisierungsleistungen eine feste Größe bei den Erschließungsarbeiten bilden.<sup>11</sup> Auffällig ist vor allem, dass die Art des Personaleinsatzes bei der Strukturdatenerschließung bzw. Tiefenerschließung ein sehr heterogenes Bild hinterlässt. In den meisten Fällen kommt für das Massen-

Strukturdatenerschließung und Projektmanagement

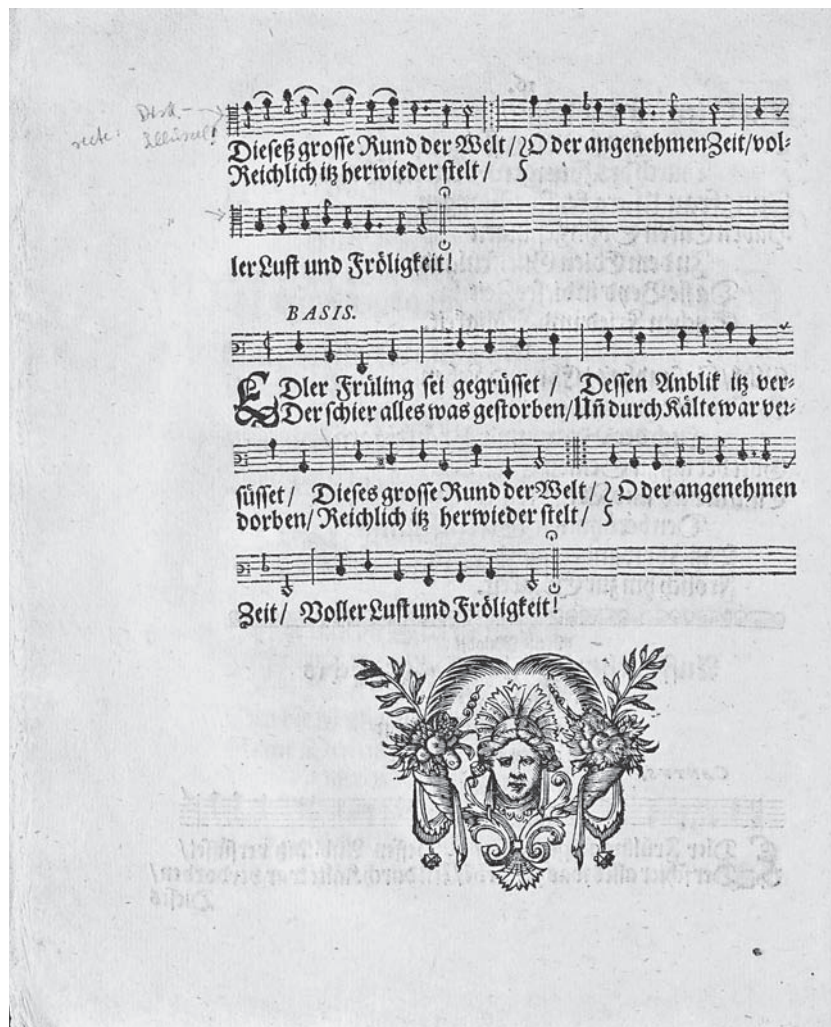


Abb. 1: Johann Rist: Hochzeitliches Frühlingsgedicht Zu sonderbaren Ehren und Gefallen Dem ... H. Henrico Stern Hochfürstlichem Braunschweigischem wolbestaltem Kantzelei Sekretarien zu Wolfenbüttel. [S.1.] 1649. HAB: Db 4505 (3)

Foto: HAB

	Drucke (Soll/Haben)	Seiten (Soll/Haben)	Aufnahmen (Haben)	Dauer/ Monate	Einseitig	Kosten/ Seite
SLUB Dresden: Digitale Quellensammlung zur Technikgeschichte. [abgeschlossen.] <sup>1</sup>	805/ 1.289	167.000/ 262.154	131.029	23	0 %	0,48 (DL)
SUB Göttingen: Digitalisierung VD17-Nova	3.200/ 300	560.000/ 40.000	24.000	24	20 %	Soll <sup>2</sup> : 0,19 € (DL)
ULB Halle: Digitalisierung der Drucke des 17. Jahrhunderts aus der Sammlung Ponickau <sup>3</sup>	9.621/ 10.187	585.000/ 582.048	ca. 292.000	24	<2 %	0,33 € (DL)
ThULB Jena: Bibliotheca Electoralis <sup>4</sup> .	2.000/ 450	520.000/ 210.000	210.000	24	100 %	120 PAM (BAT VIII = 280.000 €) Soll: 0,54 €
BSB München Digitalisierung, Erschließung, Bereitstellung im WWW und Langzeitarchivierung der Inkunabeln der BSB <sup>5</sup>	9.060/ 1.090	1.665.000/ 270.864	ca. 255.000	48	95 %	84 PAM <sup>6</sup> (BAT VII = 243.810 €) Soll: 0,21 €
BSB München: Erschließung und Bereitstellung im WWW von im deutschen Sprachgebiet erschienenen Drucken des 16. Jahrhunderts der Bayerischen Staatsbibliothek (VD 16-1) [abgeschlossen] <sup>7</sup>	4.172/ 4.234	840.000/ 580.058	ca. 495.000	24	70 %	72 PAM (BAT IXb/ VII = ca. 192.000 €) 0,33 €
BSB München: Erschließung und Bereitstellung im WWW von im deutschen Sprachgebiet erschienenen Drucken des 16. Jahrhunderts der Bayerischen Staatsbibliothek (16-2) <sup>8</sup>	36.150/ 9.182	7.230.000/ 1.348.585	ca. 890.000	24	12 %	96 PAM (BAT VII = 278.640 €) Soll: 0,04 € faktisch <sup>9</sup> ~ 0,13–0,15 €
HAB Wolfenbüttel: dünnhaupt digital – Digitalisierung der in Dünnhaupts Personalbibliographie nachgewiesenen Drucke unter Einschluss der Bibliographie von Fürstenwald/Woods <sup>10</sup>	2.000/ 1.844	532.000/ 380.490	ca. 350.000	24	88 %	48 PAM (TV L 3 = 125.972 €) Soll: 0,24 € faktisch ~ 0,26 €

Abb. 2: Übersicht zur Imagedigitalisierung. – Erl.: Die Kosten wurden anhand der bewilligten Mittel berechnet. Dabei bedeutet »Soll« die im Projektantrag genannte Zahl der zu digitalisierenden Drucke, »Haben«, der Projektstand Anfang 2009, »Aufnahmen« der »Klick« zur Erzeugung von zwei Images oder bei Einseitigkeit nur ein Image, ein »Image« bzw. Bilddatei entspricht grundsätzlich einer Seite, die Angabe »Dauer/Monate« bezieht sich auf die geplante Gesamtlaufzeit des Projekts, »Kosten/Seite« beziffert die anhand der Zahl der Aufnahmen in Relation zur nötigen Einseitigkeit errechnete reinen Scan-Kosten pro erzeugtem Image bzw. Seite (dabei werden nur die bewilligten Projektmittel berücksichtigt), DL steht für Dienstleister.

geschäft eine Mischung von bibliothekarisch geschultem Personal und Studenten zum Einsatz oder aber eine Mischung von durch Dienstleister automatisiert und händisch erfassten Strukturdaten. Wissenschaftliches Personal (Inkunabelerschließung) ist eher die Ausnahme, in dem besagten Fall geht es zudem we-

niger um direkte Digitalisierungsaufgaben als um die Kombination der Digitalisierung mit einem weiterführenden Erschließungsprojekt durch eine kunsthistorische Bilderschließung. Die konservatorische Prüfung findet, wo möglich, durch Restauratoren, in anderen Fällen durch bibliothekarisches Fachpersonal statt. Der Aufwand für die konservatorische Prüfung muss projektbegleitend gewährleistet sein und schwankt in Abhängigkeit vom Material. In der HAB Wolfenbüttel liegt er z. B. beim Projekt »Dünnhaupt digital« bei etwa einem Personenarbeitsmonat TV-L 9 pro 1000 Bänden und wird als Eigenleistung erbracht.

## Scannen und Technik

Markus Brantl und Irmhild Schäfer (BSB München) hielten das Initiativreferat zu »Massendigitalisierung: Materialität und Scantechnologie am Beispiel der Drucke des 16. Jahrhunderts«. Der Vortrag fokussierte auf die beiden Prozesse der Vorbereitung – hier vor allem die konservatorischen Vorgaben am Beispiel der BSB – und das Scannen als Durchsatzfaktor bei Altbestandsprojekten. Brantl berichtete, dass im ersten Massendigitalisierungsprojekt VD16-1 (1500–1517) 4.234 Titel digitalisiert wurden, gearbeitet wurde überwiegend mit Studenten, die in Handauflage scanneten. Das VD16-2 (1518–1600) zählt mit bisher 9.182 digitalisierten Drucken (Stand zur Zeit des Workshops) zu den ehrgeizigsten Vorhaben im Bereich der Massendigitalisierung. Hier wurde erstmals in Deutschland in größerem Umfang Robotertechnik für Altbestandsdigitalisierung eingesetzt.

Schäfer erläuterte die grundlegenden konservatorischen Vorgaben an der BSB, wie den konservatorisch guten Öffnungswinkel zur Schonung des Buchrückens und den vollständigen Verzicht auf die Glasplatte. Vor jeder Digitalisierung erfolgt grundsätzlich eine konservatorische Prüfung. Dabei ist auf die unterschiedlichen Vorlagen und deren Materialität Rücksicht zu nehmen. Alte Drucke weisen vielfältige Besonderheiten auf, wie z. B. Verwerfungen im Papier, steife Papiere, Besonderheiten der Heftung, der Bünde, der Rückenbleimung usw., die besondere Verfahren beim Scannen erfordern. Dabei sollen sich die Scanner den Büchern und nicht umgekehrt die Bücher den Scannern anpassen. So ergibt sich ein Bedarf an unterschiedlichen, für den jeweiligen Aufnahmestück optimierten Geräten und Vorlagenhalterungen. Grundsätzlich gilt an der BSB, dass ohne Glasplatte gescannt wird.

Den Einsatz der unterschiedlichen Aufnahmetechnik erläuterte Brantl am Beispiel der Hardware, die für die Digitalisierung des VD16-Projektes verwendet wurde – Handauflagescanner und Scanrobotik. Die Hardware im Handauflagebereich ist technisch auf dem Stand des Jahres 2005. So konnten auf den Handauflagescannern mit der Standard-Buchwippe, die eine 180° Öffnung des Buches voraussetzt, ungefähr 30 % der Bücher des 16. Jahrhunderts bearbeitet werden. Hinsichtlich des Durchsatzes bietet der 180° Öffnungswinkel die Möglichkeit, mit einem Scan zwei Seiten/Imagedateien abzuspeichern. Rund 70 % der Werke ließen sich nur bei Öffnungswinkeln von 90° bis 140° aufnehmen, mit der Folge, dass *verso*- und *recto*-Seiten nur getrennt digitalisiert werden konnten, d. h. ein Scan generierte nur eine Seite/Imagedatei.

Zur Hebung des Durchsatzes wurde 2007 auf eine neu entwickelte Scanrobotertechnologie umgestellt.

Die zweiseitig digitalisierenden Scanroboter arbeiten mit einem Öffnungswinkel von nur 60°. Der Durchsatz aus den VD16-Projekten wurde über einen längeren Zeitraum aus den Aufwänden für die folgenden Prozesse berechnet: Preprocessing, Scan-Operating, Postprocessing und Langzeitarchivierung der digitalen Masterdateien. Brantl erläuterte Scantechnik und Auswirkungen auf den Durchsatz auf die o. g. Teilprozesse in einer Grafik am Beispiel des Scannens in Handauflage (hier ist z. B. der Prozess separaten Scannens von *verso*- und *recto*-Seiten und deren Zusammenfügung von Zeitaufwand und Fehlerträchtigkeit her kritisch) im Vergleich zur Scanrobotik (hier fehlt z. B. noch eine Lösung für eine rasche Integration der Buchdeckelscans). Dabei wurden folgende Durchsätze erzielt: mit Handauflage werden 380 Seiten pro Tag (ein Arbeitstag entspricht acht Stunden, reale Arbeitszeit sechs Stunden, gearbeitet wird in Vier-Stunden-Schichten), am Scanrobot 1000 Seiten am Tag erreicht. Allerdings sind nur 50 % roboterfähiges Material. Zur Hebung des Durchsatzes werden seitens des Münchner Digitalisierungszentrums (MDZ) in Public Private Partnerships neue Scantechniken entwickelt, so z. B. eine neue Buchwiege für den Scanrobot, die einen frei definierten Wiegenwinkel erlauben soll. Interessant könnte auch der Einsatz eines neuentwickelten Selbstbedienungsscanners mit einer Tiefenschärfe von 19 cm sein, der es erlauben würde, auch Drucke mit sehr engen Bundstegen noch ohne Textverlust zu digitalisieren. Dieser soll in Zusammenarbeit mit dem MDZ für die Anforderungen der Altbestandsdigitalisierung angepasst werden. Zusammengefasst: Unterschiedliche Drucke müssen mit unterschiedlicher Scantechnologie aufgenommen werden, die eine Vielzahl buchtechnischer Parameter berücksichtigen. Zum Schutz der Bücher ist eine konservatorische Prüfung nötig, wichtig ist auch, dass die Restauratoren in die Entwicklung neuer Techniken mit einbezogen werden. Brantl brachte es auf die griffige Formel: Konservatorische Vorgaben + Materialität + Scantechnologie + Workflow + Personal = Durchsatz.

In der sich anschließenden Diskussion wurde bei der Bewertung der von der BSB genannten Leistungszahlen deutlich, dass die angegebenen Werte und Durchsätze von den verschiedenen Projekten durchaus nicht einheitlich erhoben werden. Um vergleichbare Zahlen zu erhalten, ist es z. B. wichtig, festzulegen, ob man den reinen Scanvorgang betrachtet oder die Rüstzeiten mit einbezieht. Gesondert zu berücksichtigen ist die Erschließung. Wenn man diese Punkte beachtet, zeigt sich, dass die Werte bei den unterschiedlichen Projekten nicht erheblich differieren. Scanzeiten sind abhängig vom Material und der verwendeten

**Massendigitalisierung von Drucken des 16. Jahrhunderts**

**380 Seiten am Tag mit Handauflage**

**1000 Seiten am Tag mit Scanrobot**

**Eine konservatorische Prüfung ist zum Schutz der Bücher nötig.**

Bei Büchern mit starker Vorschädigung ist weiterer Substanzverlust vorprogrammiert.

Technik. So erklärt sich, wie es trotz Robotertechnik zu vergleichsweise niedrigen Durchsätzen kommen kann. Bei geeignetem Material können deutlich höhere Werte erzielt werden, z. B. erreicht man bei Drucken des 19. Jahrhunderts am Scanrobot etwa 1200–1400 Seiten pro Stunde als Einzelleistung, bei denen des 17. Jahrhunderts auf dem Grazer Buchtisch jedoch nur ca. 300 Seiten pro Stunde (reine Scanzeiten).

Für Massendigitalisierungsprojekte stellt sich vor allem in der Planungsphase das grundsätzliche Problem, dass es kaum möglich ist, vorher (auch nicht mit Stichproben) abzusehen, wie das Material genau beschaffen ist. Die gleichen Ausgaben können in verschiedenen Bibliotheken je nach Einband ganz unterschiedlich beschaffen sein, sodass kaum generalisierbare Aussagen zur Digitalisierungsfähigkeit, sondern allenfalls bestandsbezogene Aussagen zum Erhaltungszustand des jeweiligen Druckexemplars möglich sind. So ist es mit Blick auf die Produktionseffizienz oft schwer einzuschätzen, was einseitig und was doppelseitig zu digitalisieren ist.

Öffnung des Buchblocks

Es zeigte sich, dass die Anforderungen an die Digitalisierung von alten Drucken material- und technikbedingt sehr komplex sind und eine flexible Herangehensweise erfordern. Je nach Material müssen verschiedene Geräte eingesetzt werden. Manches wird auf dem Roboter möglich sein, manches nur in Handauflage und hier auch, differenziert nach konservatorisch möglichem Öffnungswinkel und Falztiefe, entweder einseitig oder zweiseitig.

#### **Konservatorische und restauratorische Aspekte**

Almuth Corbach (HAB Wolfenbüttel) griff in ihrem Beitrag »Der Beitrag des Restaurators zur Massendigitalisierung – Überlegungen zur Bestandserhaltung« die bestandserhaltenden Aspekte bei der Massendigitalisierung auf und ging gezielt auf zwei Vorschläge ein, die als Reaktion auf relativ hohe Ablehnungsquoten aus konservatorischen Gründen im Projekt *dünnhaupt digital* an die HAB herangetragen worden waren: Zum einen das vorsichtige Lösen und später neue Herstellen der Bindung, zum anderen die Anwendung von technisch weniger anspruchsvollen Aufnahmeverfahren. Das Problem ergibt sich durch zahlreiche, vor allem »kleine und dicke« Objekte mit enger Bindung und extrem schmalen Bundsteg. Sie erläuterte zunächst die Kriterien der Prüfung. Problematisch für den Druck sind bei der Digitalisierung unflexible Buchgelenke oder Gelenke, die durch gebrochene Bünde oder eingerissene Überzugsmaterialien geschwächt sind, wo mit einem Ablösen der Buchdeckel oder einem Lockern bzw. Reißen der Heftung zu rechnen ist. Hohle Rücken dicker Pergamentbände werden bei der Auf-

Projekt »dünnhaupt digital«

Bestandsschutz spielt wichtige Rolle

nahme deformiert und laufen Gefahr zu knicken oder zu brechen. Bei Bänden mit festem Rücken und Vergoldung kann sich durch die Belastung einer Volldigitalisierung des ganzen Bandes das stark abgebaute Leder schollenartig ablösen und auch die Vergoldung abplatzen. Bei Büchern mit starken Vorschädigungen ist weiterer, deutlicher Substanzverlust vorprogrammiert. Schließen an Holzdeckeleinbänden mit verhärteten oder angebrochenen Riemen können durch Überdehnung sehr schnell vollständig abbrechen. Kein konservatorisches, sondern vielmehr technisches Argument ist ein zu schmaler Bundsteg, doch wird dieser bei der Prüfung mit berücksichtigt. Der Wolfenbütteler Buchspiegel, der für Massenverfahren an der HAB eingesetzt wird, benötigt je nach Buch 6–9 mm Falzbreite, um ohne Textverlust arbeiten zu können.

Corbach zeigte anhand mehrerer Beispiele, was es bedeuten könnte, die Bindung dieser Drucke zu lösen. Mit einem Schnitt durch den Falz des Einbandes ließe sich der Rücken abklappen und der Buchblock weiter öffnen. Die Konsequenz wäre ein irreparabler Schaden am Einband. Zudem entstünden hohe Kosten zwischen 200 und 500 EUR für das fachgerechte Wiedereinbinden. Corbach betonte, dass es nicht Ziel einer auch bestandserhalterisch motivierten Digitalisierung sein könne, die Objekte zu beschädigen. Grundsätzlich möglich erachtete sie den zweiten Vorschlag, nämlich beim Digitalisierungsverfahren von den bisherigen Qualitätsstandards abzuweichen. Dabei bieten sich zwei Möglichkeiten an, um einen Gutteil der im Massenverfahren abgelehnten Drucke doch noch zu digitalisieren. Entweder man akzeptiert Textverlust oder versucht aus dem Blickwinkel des Lesers in das Objekt hinein zu fotografieren, nimmt also Verzerrungen in Kauf.

In der anschließenden Diskussion sprachen sich die Teilnehmer nachdrücklich dafür aus, keine Digitalisierung mit Textverlust durchzuführen. Die Entwicklung der OCR wird in den nächsten Jahren so weit fortschreiten, dass auch Frakturschriften lesbar sein werden, sodass die Vollständigkeit von Texten eine unverzichtbare Bedingung ist, um für alle technischen Weiterentwicklungen und zukünftigen wissenschaftlichen Fragestellungen gerüstet zu sein. Ausnahmen bilden allenfalls mittelalterliche Handschriften, wo sich das Interesse weniger auf den Text als auf die Schreiberhände oder Illuminationen richtet. Es wurde noch einmal deutlich gemacht, dass keine Partei, weder die digitalisierende Einrichtung, noch der Nutzer, noch der Förderer der Digitalisierungsprojekte es gutheißen würde, die Bücher zu zerstören oder zu beschädigen. Der Bestandsschutz spielte eine große und wichtige Rolle. Bibliotheken, gerade bei Massendigitali-

lisierungsprojekten, sollten, so die Anregung, die Kooperation mit anderen digitalisierenden Bibliotheken suchen, um auf andere Druckexemplare ausweichen zu können, die sich als Vorlage besser eignen. Auf keinen Fall dürfe Bestandserhaltung gegen Digitalisierung ausgespielt werden.

Diskutiert wurde die Frage, warum die Ablehnungsrate im Dünnhaupt-Projekt der HAB bei 54 % und im VD16-Projekt in München bei 0 % liegt. Im Gespräch klärte sich, dass es unterschiedliche Ansprüche an die Scanqualität und das Verfahren gibt. Die HAB digitalisiert nur, wenn Scans entstehen, die nach gegenwärtigem Technikstand für OCR weitergenutzt werden können. Zudem werden Drucke nur berücksichtigt, wenn sie am für den Massendigitalisierungsworkflow vorgesehenen Scangerät, also am Buchspiegel, gescannt werden können. München sucht dagegen für jeden Druck die individuell beste Lösung. Kriterium für die Qualität des Digitalisates sei vor allem die Lesbarkeit. Es wurde deutlich, dass bei individueller Behandlung der Bücher vieles digitalisiert werden kann, was im normalen Workflow eines auf spezifische Verfahren ausgerichteten Massendigitalisierungsprojekts nicht möglich ist. Allerdings muss, wenn man nicht nur Masse, sondern auch Vollständigkeit anstrebt, der Zeitaufwand anders berechnet werden.

Die Runde der Teilnehmer stimmte grundsätzlich darin überein, dass die konservatorischen Probleme in den Projekten in der Planungsphase unterschätzt worden waren. Allerdings bestehen kaum Möglichkeiten, Ablehnungsquoten oder die Zahl der Fälle, bei denen einseitig gescannt werden muss, genauer als bisher vorherzusagen, da eine vorherige Gesamtsichtung der Materialien bei Massendigitalisierungsverfahren nicht durchführbar ist. Nach Maßgabe der heutigen Technik muss man bei alten Drucken mit unvorhersehbaren Problemen rechnen.<sup>12</sup> Perspektivische Lösungen liegen in der Verbesserung der technischen Infrastruktur und der fotografischen Aufnahmeverfahren, andererseits aber auch in einer stärker dezentral organisierten Digitalisierung (s. Abschnitt Kooperative Digitalisierung). Eine Beschädigung der Objekte zum Zwecke der Digitalisierung wird prinzipiell abgelehnt. Ggf. muss man Materialien zurückstellen, bis geeignete technische Verfahren gefunden sind.

**Interoperabilität, strukturelle Metadaten und persistentes Zitieren**

Dorothea Sommer (ULB Halle) führte mit dem Beitrag »Automatisierung, Qualitätssicherung, Strukturdatenerfassung« in das Thema ein. Im Projekt »Digitalisierung der Drucke des 17. Jahrhunderts aus der Sammlung Ponickau« wurden ca. 10.000 Drucke im Umfang



Foto HAB

Abb. 3: Thesaurus Exoticorum. Oder eine mit Außländischen Raritäten und Geschichten Wohlversehene Schatz-Kammer. Hamburg 1688 HAB: Gv 2<sup>o</sup>26

von ca. 600.000 Seiten digitalisiert, inhaltlich strukturiert und im Web präsentiert. Es wird mit zwei Dienstleistern zusammengearbeitet, wobei ein Dienstleister die Drucke scannt und vorstrukturiert und der andere Dienstleister für die weitgehend automatisierte Aufbereitung der Digitalisate in der Webpräsentation zuständig ist. Der geforderte Durchsatz pro Monat war die Digitalisierung und Freischaltung von 411 Drucken. Die durchschnittlich im Projektverlauf erbrachte Leistung liegt bei 494 Drucken/Monat, also bei 122 %. Dieses Ergebnis konnte durch den konsequenten Einsatz von weitgehend automatisierten Teilprozessen bei der Verarbeitung und Qualitätskontrolle der Drucke erzielt werden. Automatisiert verarbeitet wurden: die Erzeugung der Metadatenformate METS, MODS, DC auf der Grundlage der bibliografischen Daten des VD17, die URN-Vergabe (ausgeweitet auf Einzelseitenbasis im Projekt URN Granular in Kooperation mit der DNB), die Einspielung der Metadaten und Digitalisate in Portale über OAI und SRU, die Erzeugung der Titelaufnah-

Sammlung Ponickau

Zusammenfassung Strukturtypen				
Typ (Struktur)	Gesamt	Links	Freigegeben	Nicht freigegeben
Vd Vorderdeckel	5958	0	5947	11
Rd Rückdeckel	6845	0	6830	15
Inhaltsverzeichnis	190	0	190	0
Rg Register	604	0	602	2
Ab Abschnitt	70316	0	70110	206
Ti Titelblatt	9436	0	9422	14
Vo Vorwort	1911	0	1907	4
Widmung	3003	0	2997	6
Vb Vorblätter	10159	0	10146	13
Er Errata	258	0	257	1
Ex Exlibris	185	0	184	1
Ku Kupfertitel	63	0	63	0
Ba Besitznachweis	2	0	2	0
Lv Literaturverzeichnis	12	0	12	0
Kl Klassifikation	1	0	0	1
Ei Eintrag	317	0	314	3
Annotation (hs.)	7	0	7	0
Widmung (hs.)	3	0	3	0
II Illustrationen	120	0	120	0
Gedicht	24	0	24	0
Karten	8	0	8	0
<b>Σ</b>	<b>109422</b>	<b>0</b>	<b>109145</b>	<b>277</b>

Abb. 4: Häufigkeit der Vergabe von Strukturdaten im Projekt der ULB Halle

Resolving-Dienst  
der Deutschen  
Nationalbibliothek

men der Online-Ausgaben im GBV, die Verlinkung der Digitalisate im VD17. Durch laufende automatisierte und intellektuelle, qualitätssichernde Maßnahmen konnten die Fehler auf ca. 1 % reduziert werden. Bei der Strukturdatenerfassung orientiert sich die Terminologie an den in den Diskussionen zum DFG-Viewer festgelegten Begriffen. Für ca. 50 % der weniger umfangreichen Drucke genügte eine Minimalerschließung, der übrige Teil wurde durch eine bibliothekarische Fachkraft erschlossen, die zu den Strukturdaten auch noch die einzelnen Überschriften aus Kapiteln, Abschnitten etc. aus der Vorlage übertrug. Pro Tag konnten durchschnittlich 18 Drucke bearbeitet werden. Es wurden ca. 105.000 Strukturbegriffe vergeben (Verteilung s. Abb. 4).

Persistent Identifier

In einem zweiten Vortrag zum Thema erläuterte Kay Heiligenhaus (für die ULB Halle) das Konzept von »URN Granular: Persistent Identifier auf Seitenebene«. In Halle werden im Projekt URNs auf Seitenebene vergeben (nach dem DNB-Modell für Teilobjekte). Uniform Resource Names sind im Unterschied zu URLs persistente Identifikatoren (PI), weltweit eindeutige Bezeichner, die gewährleisten, dass digitale Objekte über ihren gesamten Lebenszyklus dauerhaft über Systemgrenzen und Systemwechsel hinweg identifizierbar und adressierbar bleiben – unabhängig vom Ort der Speicherung.

Teil des Konzepts ist ein Resolving-Dienst, der die

Namen und Adressen aller im System registrierten Objekte verwaltet und die Auflösung durchführt. Für die Bundesrepublik Deutschland, Österreich und die Schweiz betreibt die DNB einen solchen Resolving-Dienst. Die ULB Halle hat in Kooperation mit der DNB die Vergabe der URN, die bisher nur auf der Ausgabenebene von digitalen Objekten möglich war, auf die Seitenebene erweitert, wodurch es für die Nutzer möglich ist, die digitalen Publikationen nicht nur auf der Ausgabenebene, sondern auch auf der Seitenebene zuverlässig zu zitieren.

In der Praxis werden die Daten via OAI-pmh-Harvesting im *xepicur*-Format zweimal täglich an die DNB geschickt. Eine Konsistenzprüfung erfolgt über, von der DNB zur Verfügung gestellte, Schnittstellen. Die Erfahrungen mit dem Verfahren, bei dem inzwischen über 560.000 URNs in einem automatisierten Verfahren vergeben wurden, wurden als gut eingeschätzt. Über den Resolving-Dienst der DNB kann eine Redundanzprüfung erfolgen und über ein Qualitätsprüfungstool an der ULB gibt es Rückmeldungen, wie der Stand der Lieferungen und der Vergabe der URNs ist.

In der sich anschließenden Diskussion wurde grundsätzlich die Notwendigkeit von persistenten Links auf Werk- und auf Seitenebene hervorgehoben. Ein Diskussionspunkt war, wie im Falle von Änderungen im Digitalisat mit dem Persistent Identifier auf Seitenebene zu verfahren ist. Will man in vorhan-

denen Digitalisaten Änderungen vornehmen (Seite falsch, Seite fehlt, Seite doppelt) müsste wegen der verlässlichen Zitierung der Netzpublikation in Analogie zum Druck eigentlich eine Aufnahme für eine zweite korrigierte digitale Ausgabe erstellt werden. Die logische Notwendigkeit für diese weitere Aufnahme wurde grundsätzlich bejaht, jedoch sollte es pragmatische Grenzen geben, da dem Benutzer der digitalen Kopie solche Unterschiede nicht vermittelt werden können – und die dauerhafte Archivierung von fehlerhaftem sachlich nicht erstrebenswert ist. Dies betrifft Fehlerkonstellationen wie Ergänzen eines Images innerhalb eines Werkes, Verschieben eines Images innerhalb eines Werkes, Löschen oder Ersetzen von einem oder mehreren Images. Beim Wechsel von Zweiseitigkeit auf Einseitigkeit in der Imagepräsentation sollte jedoch grundsätzlich eine zweite Aufnahme für eine neue digitale Ausgabe erfolgen.

In diesem Zusammenhang kam nicht nur die Frage der persistenten Adressierung, sondern auch der Langzeitverfügbarkeit der Objekte zur Sprache. Die ULB Halle beabsichtigt, auch die Master-Tiffs an die DNB abzugeben, was den Sammelauftrag der DNB überschreitet, der sich auf Netzpublikationen beschränkt, wozu sich die DNB aber bereit erklärt hat. Auch die HAB ist seit längerem mit der DNB diesbezüglich im Gespräch, jedoch konnte bisher kein geeignetes Verfahren zur Datenübergabe gefunden werden. Hauptproblem ist der Transfer von Daten im Terabytebereich über das Internet bzw. ein webbasiertes automatisiertes Abgabeverfahren mit z. B. Zwischenkompression und die Frage der Rückführung von Daten im Verlustfall.

Zusätzlich zu der Vergabe von PIs für Images wurde angeregt auch auf der Strukturdatenebene granulare URNs zu vergeben (z. B. für die verschiedenen Kapitel). Darüber hinaus zeigte sich in der Diskussion, dass Strukturdaten einen bedeutenden Mehrwert nicht nur bei der Navigation im Dokument, sondern mit Blick auf mögliche Harvesting-Verfahren bei der Suche im Dokument darstellen. Dazu sei es jedoch nötig, Texte, Autoren, Überschriften auch zu transkribieren und nicht nur deren Vorhandensein zu kennzeichnen. In Göttingen wird pro Buch mit 25 Minuten für die Strukturdatenerschließung gerechnet. In Wolfenbüttel geht man von 70 durchgesehenen und annotierten Seiten pro Stunde aus. Halle erschließt nach Vorlageform durchschnittlich 40 Strukturbegriffe pro Stunde. Weniger umfangreiche Drucke werden nicht bearbeitet. München vergibt für alle in Handauflage gescannten Drucke zumindest ein Minimalset (Einband, Titelblatt, Imagezählung). Eine tiefer gehende Strukturdatenerschließung findet wegen der Mengen nicht statt, kann aber bedarfsorientiert nachträglich durchge-

führt werden, wobei der Einsatz automatisierter Verfahren angestrebt wird.

### Ökonomische Aspekte

Thomas Stäcker (HAB Wolfenbüttel) und Gabriele Meßmer (BSB München) hielten je ein Initiativreferat zum Thema »Was kostet die Digitalisierung?«. Stäcker berichtete vom Projekt *dünnhaupt digital*, in dem zum ersten Mal versucht worden war, den tatsächlichen Aufwand der jeweiligen Arbeitsschritte inklusive der Eigenleistungen zu messen. Mitarbeiterinnen und Mitarbeiter des Projektes waren gehalten, ihre Zeiten zu notieren. Der ermittelte Aufwand wurde gerundet auf die im Projekt hergestellten Seiten umgelegt und ist der folgenden Tabelle zu entnehmen (Abb. 6). Die Durchschnittssätze in Euro entsprechen noch dem BAT. Erläuterungsbedürftig sind vor allem zwei Positionen, zum einen die Nachkatalogisierung des Originals, zum anderen die hohen Strukturdatenkosten. Die Nachkatalogisierung ist eine Eigenleistung der HAB zur Katalogbereinigung. Im Zuge der seit dem Ende der 90er Jahre durchgeführten ABE Maßnahmen<sup>13</sup> konnten Titel des Altbestandes nur unzureichend katalogisiert werden. Da nun Drucke eindeutig mit persistenten Identifiern verbunden werden, besteht der Anspruch auf eine zuverlässige Zuordnung von Identifier und Druck, was eine autoptische Nachkatalogisierung erforderlich macht (im VD17 nachgewiesene Drucke sind hiervon jedoch weitgehend ausgenommen). Die Kosten für

Kosten der Digitalisierung

Langzeitverfügbarkeit der Objekte

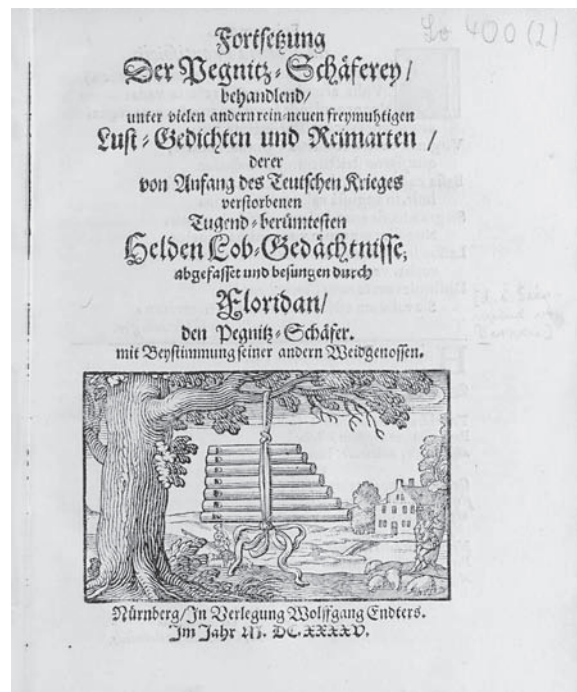


Foto HAB

Abb. 5: Sigmund von Birken: Fortsetzung Der Pegnitz-Schäferey. Nürnberg 1645. HAB: Lo 400 (2)

	Kosten/Seite	Berechnung
Imagedigitalisierung	0,24 €	Handauflage, Wolfenbütteler Buchspiegel / Aufwand für 399.003 Seiten: 36 PAMs (TV-L 3, Durchschnittssatz: € 31.783/Jahr)
Konservatorische Prüfung	0,01 €	3,08 Min/Druck für 4015 Drucke entspr. 26 Tagen, ca. 1,5 PAM TV-L 9 (Durchschnittssatz: € 43.969)
Nachkatalogisierung des Originals	0,03 €	15,11 Min/Druck für 1741 Drucke entspr. 55 Tagen, ca. 3 PAM TV-L 10 (Durchschnittssatz: € 48.257)
Katalogisierung der Sekundärausgabe	0,01 €	7,38 Min/Druck für 1561 Drucke entspr. 24 Tagen, ca. 1,5 PAM TV-L 9 (Durchschnittssatz: € 43.969)
Strukturdatenerfassung, Verlinkungen	0,40 €	Bearbeitete Seiten 87.350 722 h SHK = € 5776 + 8 PAM TV-L 9 (Durchschnittssatz: € 43.969)
Magazindienste	0,01 €	5 Min/Druck (geschätzt) für 1700 Drucke entspr. 18 Tagen, ca. 1 PAM TV-L 2 (Durchschnittssatz: € 29.883)
Auswahl, Organisation	0,04 € / 0,03 € =0,07 €	6,62Min/Druck für 4164 Drucke entspr. 57 Tagen, ca. 3 PAM TV-L 13/14 (Durchschnittssatz: € 59.698) + ca. 3 PAM (geschätzt) TV-L 9 (Durchschnittssatz: € 43.969)
<b>Summe/Seite</b>	<b>0,77 €</b>	
Weitere hier nicht berücksichtigte Kosten:		-Gerätekosten -Langzeitarchivierung -Overhead (Verwaltung, Gebäude, etc)

Abb. 6: Übersicht Kosten *dünnhaupt digital*

**Aufwand pro Buch**

Strukturdaten liegen aus der Perspektive der HAB noch zu hoch. In Zukunft kann der Anteil der bibliothekarischen Kräfte zugunsten studentischer Hilfskräfte reduziert werden. Perspektivisch wird in Folgeprojekten mit ca. 20–30 Cent/Seite gerechnet. Unklar ist nach wie vor, wie Strukturdaten kalkuliert werden müssen. Z. B. rechnet Halle in vergebenen Strukturbegriffen pro Stunde/Person und kalkuliert daraus ableitend den Aufwand pro Tag, der durchschnittlich bei 18 Drucken/Tag/Person (einschließlich der vertieften Beschreibung nach Vorlage) liegt. Göttingen berechnet den Aufwand pro Buch, der bei 25 Minuten liegt, wobei es aber den Umfang der Drucke und die Tiefe der Erschließung unberücksichtigt lässt. Mit anderen Worten, ein Buch mit 500 vergebenen Strukturinformationen zählt genauso viel wie ein Buch mit 10 vergebenen Strukturinformationen. Eine Kalkulation pro Seite wie in Wolfenbüttel lässt wiederum nicht erkennen, wie viele Daten erhoben wurden.

**ausgefeilte Geschäftsgänge sind nötig**

Meßmer legte aus den nun über drei Jahre erprobten Erfahrungen der BSB in Massendigitalisierungsprozessen detaillierte Zahlen vor. Zunächst wird noch einmal der gesamte Workflow des Scanprozesses differenziert erläutert. Diese Darstellung ist nötig, um Vergleichbarkeit bezüglich der einbezogenen Prozesse zu erreichen. Die durchschnittlichen Leistungszahlen stellen sich wie folgt dar: bei Handschriften und schwierigen Vorlagen werden 200 Seiten/Tag, bei historischem Altbestand 380 Seiten/Tag in Handauflage (70 % einseitig und 30 % zweiseitig) und bei Scannen

des historischen Altbestandes mit dem Scanrobot 1000 Seiten/Tag erreicht.

Dabei entfallen bei der Digitalisierung in Handauflage pro Druck für den gesamten Digitalisierungs-Workflow auf Preprocessing fünf Minuten, auf das Scannen von durchschnittlich 200 Seiten inklusive Einband und Spiegel (Handauflage; 70 % einseitig, 30 % zweiseitig) 250 Minuten, auf die Qualitätskontrolle und Vergabe von formalen Strukturdaten, der Freigabe, dem Prüfen der Langzeitarchivierung 30 Minuten, sodass insgesamt 285 Minuten bzw. 4,75 Stunden pro Druck benötigt werden. Umgerechnet entstehen für das Scannen Kosten von 91 €, der gesamte Prozess im Scanzentrum kostet pro Buch 103 € (ca. 0,52 € pro Seite). Die Digitalisierung mit dem Scanrobot kostet unter gleichen Voraussetzungen pro Druck nur 40 € für das Scannen und 58 € für den gesamten Scanprozess (ca. 0,29 € pro Seite).

Meßmer zog das Fazit, dass Massendigitalisierung ausgefeilte Geschäftsgänge und genügend Kapazität für die in Eigenleistung zu erbringenden Prozess-Schritte erfordere. Nötig sei ein kontinuierliches Aufrüsten der vorhandenen bzw. das Austauschen veralteter Scanner. Die Kosten sind je nach Materialart und -besonderheiten sehr unterschiedlich. Das Scannen mit Robotern ist deutlich günstiger, doch lassen sich im Altbestand (zurzeit) allenfalls 50 % der Werke am Roboter scannen. Zu erwarten ist jedoch, dass durch die Weiterentwicklung von Scannern die Scankosten weiter sinken werden.

In der Diskussion wurde noch einmal unterstrichen, dass, um möglichst kostengünstig und effizient produzieren zu können, für alle Geräte ein Belegungsplan wichtig sei. Es wurde ferner darauf aufmerksam gemacht, dass die Anschaffungskosten und Betriebskosten des Scanrobots in Relation zu den günstigeren Scankosten gesetzt werden müssten. Hinzu kommen bei jeder Digitalisierung die Betriebskosten für die Langzeitarchivierung, die inzwischen höher liegen als die Speicherkosten.

### **Kooperative Digitalisierung**

Thomas Bürger (SLUB Dresden) führte in das Thema mit einem Beitrag »Von der Personalbibliographie zur digitalen Werkausgabe. Wege zu einer deutschen digitalen Autorenbibliothek am Beispiel des Werkes von Christian Weise (1642–1708)« ein. Da die Forschung an den Universitäten zunehmend im Netz stattfindet, wird es immer wichtiger, die eigenen Bestände online zugänglich zu machen und dabei kooperativ vorzugehen. Bestandserhaltung und Digitalisierung sind zwei Seiten einer Medaille und dürfen nicht gegeneinander ausgespielt werden. Bürger demonstrierte am Beispiel der Überlieferung des Werkes von Christian Weise die Notwendigkeit zur Kooperation. Die kritische Werkausgabe zu Weise umfasst nach 38 Jahren inzwischen 16 Bände. Alle zwei Jahre ist durchschnittlich ein Band erschienen. In der Bibliographie von Dünnhaupt werden 341 zeitgenössische Ausgaben Weises verzeichnet, davon 159 Erstdrucke. In der HAB befinden sich 118 Drucke, das entspricht einem Drittel der zeitgenössischen Ausgaben. Insgesamt sind in den VD 17-Bibliotheken Wolfenbüttel, Halle, Dresden, München und Göttingen 75 % der Ausgaben vorhanden. Derzeit sind aber erst 14 % digitalisiert (Wolfenbüttel: 30, Halle 15, Dresden 1), sodass der Forschung die Quellen noch nicht in dem wünschenswerten Umfang zur Verfügung stehen.

Diese exemplarische Auswertung verdeutlicht: Die Digitalisierung ermöglicht grundsätzlich, forschungsrelevante Textcorpora in kurzem Zeitraum virtuell zusammenzuführen und damit der Forschung quantitativ und qualitativ verbesserte Arbeitsmöglichkeiten zu verschaffen. Angesichts der Streuung der Originalbestände hängen sichtbare Erfolge für die Forschung jedoch von einer kooperativen Digitalisierung in den Bibliotheken ab. Diese ist wiederum zur effektiven Steuerung auf einen zentralen Nachweis der digitalisierten Drucke angewiesen. Auch der Nutzer braucht *einen* Katalog, in dem alles (auch VD 16 und VD 17) zusammengeführt ist, mit dem Nachweis der Digitalisate. Aus Sicht der Forschung ist *ein* nationaler Katalog mit dem Nachweis der erhaltenen Originale und dem Link zu den Digitalisaten dringend geboten. Die Frage,

ob dies ein erweiterter KVK leisten kann oder sich die Verbünde – angesichts der Google-Konkurrenz – auf eine gemeinsame Suchmaske verständigen können, hat deshalb höchste Priorität.

Rubert Schaab (SUB Göttingen) forderte in seinem Anschlussvortrag »Organisatorische und technische Aspekte eines koordinierten Vorgehens« Digitalisierung solle bibliotheksübergreifend, arbeitsteilig und nach dem Prinzip der Gegenseitigkeit erfolgen. Als Organisationsmodell und Arbeitsinstrument für ein solches Vorgehen zur Digitalisierung des Altbestandes können Verfahren der GBV-Online-Fernleihe dienen, die für einen begrenzten Teilnehmerkreis aus dem GBV als eigenes Verfahren der beschleunigten Fernleihe aufgelegt werden kann. Dabei stellt die Ermittlung des für die Digitalisierung am meisten geeigneten Exemplars eine besondere Anforderung dar. Die Teilnehmer erhalten eine Digitalisierungsanforderung auf ein bestimmtes Buch, wie bei der Fernleihe wird dann nacheinander abgefragt, wer ein geeignetes Exemplar hat und es digitalisieren kann. Wichtig ist die Ausgeglichenheit von Geben und Nehmen, nach und nach kann der Teilnehmerkreis erweitert werden. Wünschenswert wäre auch ein verbundübergreifendes Verfahren, das ist jedoch zurzeit nicht realistisch.

In der sich anschließenden Diskussion wurde zu bedenken gegeben, dass schon heute bei der Fernleihe ein Problem von Geben und Nehmen besteht. Das Prinzip der Kostenfreiheit sei zwar grundsätzlich zu begrüßen, aber bei Anfall größerer Mengen kaum realistisch, da es bei Überlastung des Systems dazu führen würde, dass keine der beteiligten Bibliotheken mehr digitalisiert. Zumal sei der logistische Aufwand in diesem Modell für die avisierte Masse von Drucken (600.000 sollen es allein im VD 18 sein) erheblich und kaum zu leisten.

Es entspann sich eine intensive Diskussion um den immer wieder geforderten zentralen Nachweis von Digitalisaten in Deutschland. Derzeit werden in Deutschland Digitalisate hauptsächlich in den Verbundkatalogen und in den retrospektiven Nationalbibliografien verzeichnet. Es ist wichtig, dass die Digitalisate, die primär in den Verbänden nachgewiesen werden, auch – bei entsprechender Relevanz – an VD 16 und VD 17 gemeldet und dort in automatisierten Verfahren anhand der Normnummern nachgetragen werden. Die immer wieder angekündigte Verbundintegration, die die föderal bedingte uneinheitliche Nachweissituation überwinden sollte, ist bislang nicht erfolgt. Hinzu kommt, dass weder alle Verbünde noch der von Forschern intensiv genutzte KVK<sup>14</sup> eine gezielte Suche bzw. Selektion von Digitalisaten erlauben. Das Portal zvdd<sup>15</sup> wiederum hat bisher nicht die

**koordiniertes Vorgehen**

**verbundübergreifendes Verfahren wäre wünschenswert**

**Streuung der Originalbestände**

**Normnummer**

gesetzten Erwartungen erfüllt. Der auch Strukturdaten umfassende Relaunch der Datenbank ist bisher nicht erfolgt, auch wenn wichtige Klärungen bei der Formatfrage erfolgt sind, wie aus Göttingen zu hören war. Im Weiteren sind die Perspektiven für die Deutsche Digitale Bibliothek bisher ungewiss. Hingewiesen wurde auf EROMM,<sup>16</sup> das sich auch die Aufgabe des Nachweises von Masterdigitalisaten stellt. Auch die Deutsche Nationalbibliothek (DNB) könnte hier eine Rolle spielen, da sie den gesetzlichen Auftrag habe, Netzpublikationen, ergo auch die von den Projekten angebotenen Digitalisate, zu archivieren. Dennoch scheint unwahrscheinlich, dass seitens der DNB kurzfristig mit einem zentralen Nachweis der digitalisierten Drucke zu rechnen ist. Genannt wurde ferner OAlster,<sup>17</sup> an dem das für alte Drucke unzureichende Erschließungsniveau auf der Basis von DC kritisiert wurde. Der Mehrwert gegenüber Google sei nicht allzu groß. Dennoch sollte OAlster bedient werden, zumal es nach dem Eindruck der Teilnehmer perspektivisch vor allem OAI-Verfahren sind, die die Zukunft bestimmen werden. Die auch durch den DFG-Viewer<sup>18</sup> etablierten Standards (METS/MODS), so die Überzeugung, werden den Datenaustausch und die Datenaggregation erheblich erleichtern.

Angesichts der derzeitigen Bedeutung des KVK, den vor allem Forscher neben Google gern nutzen, sollte zunächst vor allem dort eine Möglichkeit geschaffen werden, gezielt nach Digitalisaten zu suchen. Konsens war, dass dessen ungeachtet möglichst viele der Nachweisinstrumente und -kataloge bedient werden sollten, im Zuge der Massendigitalisierungen von deutschem gedrucktem Kulturgut vor allem das VD 16, das VD 17 und perspektivisch das VD 18.

### Abschlussdiskussion

Die Abschlussdiskussion drehte sich nach einem kurzen Resümee noch einmal schwerpunktmäßig um das Thema Strukturdaten. Hervorgehoben wurde, dass manche Objekte (z. B. Lexika, Enzyklopädien) ohne Strukturdaten online nicht sinnvoll benutzbar seien, wodurch der Sinn der Digitalisierung nicht ausreichend erfüllt würde. Gleichwohl kann man sich auch Projekte vorstellen, die zunächst ohne Strukturdaten arbeiten. Hier kam auch das Verhältnis der Strukturdaten zu der nachträglichen Volltextgenerierung zur Sprache. Strukturdaten seien auch sinnvoll, wenn der Volltext vorliegt. Sie erlauben einen gezielten Zugriff auf logische Texteinheiten. Seiten mit Illustrationen, Druckermarken und Provenienzangaben können dank ihrer gezielt angesteuert werden, was durch einen Volltext nicht eo ipso gegeben ist. Gleichwohl wird es perspektivisch möglich sein, manche Struktur-

elemente automatisiert zu erkennen und zu erfassen (z. B. Abbildungen und Seitenzählungen).

Das Thema Volltextgenerierung und OCR-Genauigkeit wurde zumindest gestreift. Für alte Drucke gibt es viel versprechende Entwicklungen, allerdings sind die durch Abtippen in Asien erreichten Genauigkeiten derzeit noch nicht automatisiert realisierbar.

### Resümee

Der *Workshop* hat in einem Überblick versucht, alle relevanten praktischen Fragen rund um die Massendigitalisierung von Altbestand aufzugreifen. Angesprochen wurden Scannertechnik, konservatorische und ökonomische Aspekte sowie Einzelaspekte der Erschließung. Deutlich geworden ist aus den Pilotprojekten, dass alte Drucke auch unter dem Blickwinkel der Massendigitalisierung einer differenzierten Behandlung bedürfen. Bibliotheken, die sich zur Digitalisierung entschließen, sollten damit rechnen, dass je nach Buchbestand mehrere Verfahren zur Anwendung kommen müssen. Materialbezogen sind unterschiedliche Öffnungswinkel und spezielle Techniken erforderlich, die die Planungszielgrößen beeinflussen. Auch Dienstleistern gegenüber muss aus Gründen des Bestandschutzes auf eine angemessene Scan- und Buchhalterungstechnik Wert gelegt werden. In der Frage der notwendigen Beachtung des Bestandsschutzes waren sich alle Beteiligten einig. Ein Gutteil der zukünftigen technischen Entwicklungen wird diesem Segment gewidmet sein, um den Scandurchsatz bei optimaler Bestandsschonung zu erhöhen und auch nach wie vor bestehende Probleme der Aufnahme von Drucken mit zu schmalen Bundsteg zu bewältigen. Theoretische Überlegungen zum Ausbinden von Drucken sind angesichts der zu schützenden Integrität der historischen Quellen abzulehnen und wären außerdem mit sehr hohen Kosten verbunden – zumal in einem Massendigitalisierungsprojekt.

In der Bilanz bot der *Workshop* allen Beteiligten eine gute Gelegenheit, Erfahrungen auszutauschen und Anregungen in die eigenen Häuser mitzunehmen. Zum ersten Mal liegen auch belastbarere Zahlen zum Scanprozess und der Erschließung der Digitalisate vor. Obwohl es bestands- und technikbedingt Schwankungen gibt, sind doch die Leistungs- und Kostenzahlen bei allen Beteiligten so, dass man heute auf der Basis dieser unabhängig voneinander ermittelten Werte zuverlässiger voraussagen kann, wie arbeits- und kostenintensiv eine Digitalisierung von Material in Handauflage unterhalb von 90° Öffnungswinkel oder von robotergeeignetem Material ist. Weniger klar ist – und das liegt in der Natur der Sache – mit welchem Verteilungsschlüssel man bei bestimmten Beständen rech-

nen kann. Ob 10 % oder 90 % einseitig, ob 30 % oder 70 % mit einem Roboter zu digitalisieren sind, kann man im Vorfeld kaum zuverlässig ermitteln.

Trotzdem sind zur Ermittlung der Kosten noch weitere differenzierende Analysen erforderlich, um die dem Scanprozess inhärenten Komponenten besser unterscheiden und vergleichen zu können. Beim Aufwand für die konservatorische Prüfung sowie Erschließungsverfahren und -kosten gibt es zwar eine sich stabilisierende Tendenz, doch sind die Ergebnisse weniger eindeutig. Während die Katalogisierungsarbeiten für die Sekundärformen im engen Sinne relativ gut fassbar und mit knapp acht Minuten nach der Wolfenbütteler Ermittlung pro Druck anzusetzen sind bzw. wie in Halle durch ein automatisiertes Verfahren bei Vorhandensein qualitativ hochwertiger bibliografischer Primärdaten erzeugt werden können, weichen die Daten bei der Erschließung mit Strukturdaten teilweise erheblich voneinander ab, was auf eine relativ heterogene Vorgehensweise schließen lässt.

Neben den beeindruckenden Leistungen der Massendigitalisierungsprojekte wurden auf dem *Workshop* aber auch Desiderate sichtbar, so im zu verbessernden zentralen Nachweis von Digitalisaten und in der Koordinierung von Digitalisierungskampagnen. Auch gilt es, neben den Massendigitalisierungsprojekten stärker an Nutzerwünschen orientierte Digitalisierungen umzusetzen. Verfahren wie *digitization on demand* oder *digitization on use* wären sinnvollerweise in die Massendigitalisierungsprozesse zu integrieren, die hier entstehenden Digitalisate genauso in den Verbänden wie in den retrospektiven Nationalbibliografien nachzuweisen.

Trotz aller Probleme wurden wichtige Hürden genommen und prototypische Verfahren in der Massendigitalisierung etabliert. Die DFG hat, das wurde im *Workshop* ausdrücklich gewürdigt, gerade durch die Vorgabe der Praxisregeln entscheidend zur Standardisierung und durch die Förderung der Projekte zur nachhaltigen Verbesserung des Angebots an Digitalisaten in Deutschland beigetragen. Der Austausch der Praktiker, auch dies war Konsens, sollte fortgesetzt werden.

<sup>1</sup> <http://technikgeschichte.slub-dresden.de/projekte/>

<sup>2</sup> Mit den Soll-Werten sind die im Projektantrag zugrunde gelegten Mengen gemeint. Es sind keine Aussagen darüber, ob diese Werte auch erreicht wurden.

<sup>3</sup> <http://digitale.bibliothek.uni-halle.de/>

<sup>4</sup> [www.thulb.uni-jena.de/DFG\\_Projekte.html](http://www.thulb.uni-jena.de/DFG_Projekte.html)

<sup>5</sup> [www.bsb-muenchen.de/Inkunabeln.181.o.html](http://www.bsb-muenchen.de/Inkunabeln.181.o.html)

<sup>6</sup> Für die erste Bewilligungsphase von 24 Monaten, beantragt waren 4,5 Stellen für 2 Jahre.

<sup>7</sup> [www.digitale-sammlungen.de/index.html?c=sammlungen&l=de](http://www.digitale-sammlungen.de/index.html?c=sammlungen&l=de)

<sup>8</sup> [www.digitale-sammlungen.de/index.html?c=sammlungen&l=de](http://www.digitale-sammlungen.de/index.html?c=sammlungen&l=de)

<sup>9</sup> Faktisch bedeutet: die nach dem bisherigen Stand (31.1.2009) ermittelten faktischen Kosten. Diese können sich im Laufe des Projektes noch ändern.

<sup>10</sup> <http://diglib.hab.de/?link=017>

<sup>11</sup> Ausnahme ist hier die ULB Halle, die in Zusammenarbeit mit ihrem Dienstleister und dem GBV im Rahmen ihres Projekts für die Generierung der Sekundärformen ein automatisiertes Verfahren entwickelt hat. Voraussetzung hierfür sind die bereits vorliegenden hochwertigen bibliographischen Daten aus dem VD17. So konnten alle 10.187 Titel der Sekundärformen in einem automatisierten Verfahren generiert und in den GBV eingespielt werden.

<sup>12</sup> An der HAB Wolfenbüttel und an der Staatsbibliothek zu Berlin sind Größenordnungen von 20–30 % Problemfällen aufgetreten, d. h. Fälle, die ggf. einer Sonderbehandlung bedürfen und nicht im regulären Massenworkflow bearbeitet werden können.

<sup>13</sup> Haller, Klaus. Altbestandserschließung in wissenschaftlichen Bibliotheken: ein Förderprogramm der Deutschen Forschungsgemeinschaft. Berlin: Dt. Bibliotheksinst., 1995.

<sup>14</sup> [www.ubka.uni-karlsruhe.de/kvk.html](http://www.ubka.uni-karlsruhe.de/kvk.html)

<sup>15</sup> [www.zvdd.de](http://www.zvdd.de)

<sup>16</sup> [www.eromm.org/](http://www.eromm.org/)

<sup>17</sup> [www.oaister.org/](http://www.oaister.org/)

<sup>18</sup> <http://dfg-viewer.de/>

**differenzierte Analysen  
zur Kostenermittlung**

## DIE VERFASSER

**Andrea Opitz**, Arbeitsgruppe Wolfenbütteler Digitale Bibliothek, Herzog August Bibliothek, Postfach 1364, 38299 Wolfenbüttel

Tel.: +49-5331-808-314

[opitz@hab.de](mailto:opitz@hab.de)

**Dr. Thomas Stäcker**, Leiter Abteilung Alte Drucke, Digitalisierung, Herzog August Bibliothek, Postfach 1364, 38299 Wolfenbüttel

Tel.: +49-5331-808-119

[staecker@hab.de](mailto:staecker@hab.de)

*DBSM*

**IN NEUEM GEWAND ... DAS DEUTSCHE  
BUCH- UND SCHRIFTMUSEUM DER  
DEUTSCHEN NATIONALBIBLIOTHEK  
WIRD 125 JAHRE ALT UND ERHÄLT IN  
LEIPZIG EIN NEUES DOMIZIL**

### Prolog: Aufbruch

Das Deutsche Buch- und Schriftmuseum der Deutschen Nationalbibliothek hat Anlass zum Feiern: Es wird in diesem Jahr 125 Jahre alt. 1884 als Deutsches Buchgewerbemuseum in Leipzig gegründet, gehört es zu den weltweit ältesten und nach Umfang und Qualität der Bestände bedeutendsten Museen auf dem Gebiet der Buchkultur. Aber es feiert nicht nur Geburtstag, es erhält im vierten Erweiterungsbau der Deutschen Nationalbibliothek großzügige Ausstellungsflächen und Magazine, einen Speziallesesaal und neue

**Deutsches Buchgewerbemuseum 1884 gegründet**