



Digitalisierung und Langzeitarchivierung von Fotonegativen

Ein Erfahrungsbericht aus dem Kunsthistorischen Institut in Florenz – Max-Planck-Institut

Christina Bieber, Werner Schweibenz

Kunsthistorisches Institut in Florenz
Max-Planck-Institut
Via Giuseppe Giusti 44
I-50121 Firenze
bieber@khi.fi.it, schweibenz@khi.fi.it

Zusammenfassung

Der Beitrag beschreibt die Erfahrungen der *Fotothek des Kunsthistorischen Instituts in Florenz – Max-Planck-Institut* bei der Digitalisierung und Langzeitarchivierung von Fotonegativbeständen im Rahmen eines DFG-Projekts. Als Schwerpunkte werden behandelt: die Gründe für die Digitalisierung und Langzeitarchivierung der Bestände, Fragen des Retrievals und der Präsentation der Inhalte der Bilddatenbank und *digilib*, die Digitalisierung der Bestände sowie die noch offenen Fragen der Langzeitarchivierung.

Abstract

The paper provides a summary of the experiences of the photo library of the German Art History Institute (Max-Planck-Institute) in digitizing and archiving photo negatives. The focus is on reasons for digitization and long-term archiving, information retrieval and presentation of image content using *digilib*, how the images were digitized and which questions remain for long-term archiving.

1 Kurzvorstellung KHI

Das *Kunsthistorische Institut in Florenz* (KHI) blickt auf eine über hundertjährige Geschichte zurück. 1897 wurde es als eine private Einrichtung durch einen Kreis von Kunsthistorikern mit dem Ziel gegründet, die



Erforschung der italienischen Kunstgeschichte zu unterstützen. Zu diesem Zweck werden eigene Forschungsprojekte durchgeführt, regelmäßige Veranstaltungen organisiert wie z. B. Vorträge, Kolloquien, Fachtagungen oder Studienkurse für den wissenschaftlichen Nachwuchs und daraus resultierende Ergebnisse publiziert. Bibliothek und Fotothek dienen Wissenschaftlern jeder Nation als Informationsbasis für Forschungsprojekte. Seit 2002 befindet es sich das KHI in der Trägerschaft der *Max-Planck-Gesellschaft zur Förderung der Wissenschaften e.V.* (MPG) und ist somit neben der *Bibliotheca Hertziana* in Rom das zweite Max-Planck-Institut (MPI) zur Italienforschung. Das Institut zählt rund 60 Mitarbeiter und 10 Stipendiaten.

Das zu beschreibende Projekt ist an der Fotothek des KHI angesiedelt. Hier können ca. 600.000 Fotografien zu Kunstwerken Italiens vom Mittelalter bis zur Gegenwart konsultiert werden. Die besondere Qualität für die Forschung liegt in der allgemeinen Breite und Vollständigkeit des Bestandes mit z. T. umfangreichen, geschlossenen visuellen Dokumentationen zu einzelnen Kulturgütern. Der Umfang und die Dichte der kontinuierlich aufgebauten Bestände erlaubt es oft, Veränderungen über einen längeren Zeitraum systematisch zurückzuverfolgen, und dies teils bis zu den Anfängen der Fotografie.

2 Gründe für eine Digitalisierung und Langzeitarchivierung

Bis 1993 erfolgte die Erschließung des Fotothekbestandes konventionell über Zettelkataloge. Seitdem werden Neuzugänge und dazugehörige Altbestände mit der Software *HiDA* wissenschaftlich dokumentiert. Bisher konnten Metadaten zu ca. 15% des Bestandes eingearbeitet werden. Eine Recherche ist sowohl über die Web-Site¹ des Instituts möglich als auch über die vom *Bildarchiv Foto Marburg*² koordinierte Verbunddatenbank, in die die Datenbestände fast aller großen deutschen Kunstmuseen, Denkmalpflegeämter und weiterer wissenschaftlicher Institute einfließen. Von Seiten des KHI wurden in die Datenbank bisher lediglich Textinformationen zu den Fotografien angeboten, nicht jedoch die Bilder selbst. Eine Konsultation des gesamten Bestandes erfordert folglich eine Reise nach Italien und somit auch ausreichende Zeit und Geldmittel sowie einen Nachweis der wissenschaftlichen Zielsetzung, um Zugang zum Institut zu erhalten.

¹ Deutsches Kunsthistorisches Institut in Florenz. URL = <http://www.khi.fi.it>.

² Bildarchiv Foto Marburg. URL = <http://www.fotomr.uni-marburg.de>.

Durch die technischen Möglichkeiten ist es mittlerweile eine selbstverständliche und völlig gerechtfertigte Anforderung seitens der Nutzer (und Geldgeber), die Bildmaterialien selbst über das Internet zugänglich zu machen und weitere bestehenden Möglichkeiten (Mehrwerte) des Netzes sinnvoll zu nutzen. Hier galt es Anschluß zu finden an Entwicklungen und Angebote, die in anderen - vor allem kommerziellen Bereichen - längst üblich sind. Angesichts zunehmender kommerzieller und auch monopolistischer Tendenzen beim Zugang insbesondere zum fotografischen kulturellen Erbe geschieht dies am KHI bewusst mit einer nicht-kommerziellen Zielsetzung in der Absicht, alternative Strukturen zu Forschungszwecken zu erarbeiten. Das benötigte technische Know-how rund um den Aufbau, die Pflege und Verwaltung eines digitalen Bildarchivs wird daher langfristig innerhalb der Einrichtung auf- und ausgebaut, um damit auch den Wechsel von der konventionellen auf die digitale Fotografie und die damit verbundenen Arbeitsprozesse vorzubereiten.

In Anbetracht einer erhöhten Nutzung der Bildmaterialien sollte gleichzeitig einerseits die Schonung der analogen Originalnegative sichergestellt sein (scan once, use many), während andererseits Vorkehrungen zur langfristigen Sicherung der Digitalisate getroffen werden.

3 Projektvorstellung

Die Pionierphase konnte im Rahmen eines zweijährigen Projektes bis zum September 2004 realisiert werden, für das die *Deutsche Forschungsgemeinschaft (DFG)* die Kosten für die Personalmittel übernahm und zwar sowohl für die Digitalisierung als auch für die wissenschaftliche Dokumentation von 20.000 Aufnahmen. Hierfür konnten eine promovierte Ganztagskraft, zwei Halbtagskräfte und eine studentische Hilfskraft unter Vertrag genommen werden. Gelder für drei Arbeitsplätze wurden ebenfalls von der DFG bereit gestellt. Ein wesentlicher Aspekt des Projekts ist die Erstellung einer Web-Präsentation, bei der Bild- sowie Textinformationen über eine neu zu gestaltende Oberfläche mit verbessertem Retrieval und einem leistungsfähigen Bildviewer gemeinsam recherchierbar und wissenschaftlich nutzbar gemacht werden. Für Nutzer mit weitergehenden Interessen ist zudem eine Bestellfunktion geplant. Auf einige kritische Aspekte der Datenbankstruktur für das Information Retrieval soll hier eingegangen werden, ebenso auf einige ausgewählte Fragen der Gestaltung der Bilddatenbank.

Im Rahmen des Projekts gewann auch der Aspekt des Urheberrechts an Fototheksmaterialien neue Bedeutung. Bisher war es für viele Fototheken nur von Bedeutung, überhaupt einen Bildnachweis zu einem bestimmten Objekt zu haben. Mit der Digitalisierung und der Verbreitung von Fototheksmaterialien über das Internet, stellt sich die Frage des Urheberrechts am Bildmaterial bzw. den Negativen. Das KHI besitzt durch Ankäufe und Schenkungen sowie eigene Fotokampagnen zur gezielten Bestandsabrundung und -vergrößerung das Urheberrecht an ca. 50.000 Fotografien bzw. Negativen.

Für die Digitalisierung wurden die wichtigsten Originalbestände (also Aufnahmen, bei den das KHI über das Urheberrecht verfügt) ausgewählt, beispielsweise

- das Archivio Lotz mit Aufnahmen zur Florentiner Urbanistik vor dem zweiten Weltkrieg,
- die geschlossene fotografische Dokumentation der Freskenausstattung der Kirchen in Assisi aus den 70er Jahren (diese bildeten eine wichtige Grundlage für die Restaurierungsarbeiten nach dem Erdbeben im Jahre 1997),
- Architektur und Ausstattung von Sieneser Kirchen, oder
- die Ausmalung Florentiner Paläste.

4 Problemfeld Information Retrieval und Bilddatenbank

Für die Erfassung und Pflege der Fotodaten wird in der Fotothek seit 1993 die Datenbank HiDA (*Hierarchischer Dokument-Administrator* in der Version 3.0) in Kombination mit dem kunsthistorischen Inventarisierungssystem *Marburger Informations-, Dokumentations- und Administrations-System (MIDAS)* verwendet (Bove, Jens/Heusinger, Lutz/Kailus, Angela 2001) verwendet. HiDA3 ist ein etwas in die Jahre gekommenes hierarchisch organisiertes Datenbanksystem mit graphischer Benutzeroberfläche, das im Verbund des *Digitalen Informationssystems für Kunst- und Sozialgeschichte (DISKUS)* eingesetzt wird, dem auch die Fotothek angehört. Aufgrund der Zugehörigkeit zu diesem Verbund müssen einige Nachteile von HiDA in Kauf genommen werden, beispielsweise, dass HiDA in der Version 3.0 nur in sehr eingeschränktem Umfang Bilddateien verwalten und einbinden kann. Die Software ermöglicht eine sehr dezidierte, wissenschaftliche Beschreibung von kulturhistorischen Objekten vielfältigster Art, doch gerade komplexe Suchfragen führen nicht immer zu befriedigenden Ergebnissen. Es wurde daher die

Entscheidung getroffen, HiDA zwar weiterhin für die Eingabe und die Zusammenarbeit im Verbund zu nutzen, aber für ein verbessertes Retrieval sowie kombinierte Text-/Bild-Darstellung nach neuen Möglichkeiten zu suchen. Die Wahl fiel auf das als Open-Source-Produkt bestehende Content-Management System *ZOPE/Plone*.

Hierbei wird eine Datenausgabedatei in HiDA erzeugt. Die hierarchischen Strukturen werden über Skripte zerlegt, umgestaltet und in eine relationale Datenbank geladen. Die Festlegung und retrievaldienliche Steuerung dieser Prozess der Strukturauflösung, Umgestaltung sowie neuen Verknüpfung und Indexierung war mit einigem Aufwand verbunden. Ein weitere Konsequenz waren intensive Datenpflege und die Überarbeitung einiger Eingaberegeln zu Gunsten eines verbesserten Retrievals.

4.1 Indizes und Textdarstellung mit ZOPE

Das Retrieval in der Bilddatenbank erfolgt über verschiedene Indizes, die auf den entsprechenden Angaben aus HiDA basieren und für die Ausgabe im Web visuell aufbereitet wurden. Die wichtigsten sind:

- **Künstlerindex:** er bietet einen Überblick über alle in der Datenbank vorhandenen Künstler, in ihrer lexikalisch üblichen Form (Thieme-Becker/Allgemeines Künstlerlexikon).
- **Ortsindex:** eine alphabetische Liste aller in der Datenbank vorhandenen Ortsnamen oder die Suche über eine differenzierte Liste nach Regionen in Italien. Jeder Ortslink führt zu einer Liste aller erfassten Bauwerke und Monumente.
- **Gesamtindex:** erlaubt unter anderem die Suche nach Entstehungsort, Datierung, Gattung, Standort, Bildinhalt.

Ferner ist es möglich auch nach folgenden Kriterien direkt zu suchen: dargestellte Personen, Fotografen, sonstige Personen (hier sind z.B. Auftraggeber oder Besitzer gemeint), alle Personen (Gesamtindex Personen), Künstlernationalität, Material / Technik, Standort, Art des Werks, Bauwerk oder Bildinhalt (hier finden sich die Hauptkategorien der *Iconclass*-Thesaurus).

Für die Indexsuche gelten folgende **Grundregeln**:

- Mit Hilfe der Listenindizes (erreichbar über den Button „Index“) kann mit einem Phrasen- bzw. einem **Keywordindex** gesucht werden, bei selbständiger Eingabe wird im Einzelwortindex gesucht (**Freitextsuche**).
- **Abkürzungszeichen** ist der Stern: "corregg*" (anstatt Correggio).
- Suchbegriffe können durch "**and**" oder "**or**" miteinander verknüpft werden. Werden die Begriffe einfach aneinander gereiht, gilt automatisch die Verknüpfung mit „und“, z.B. „florenz assisi“.
- **Umlaute** werden als Umlaute eingegeben.
- **Apostrophs** müssen weggelassen werden: "Abate, Nicolo dell" statt "Abate, Nicolò dell' ".

Verknüpfungen mit „und“ sind besonders dort sinnvoll, wo verschiedene Felder in einem Index gebündelt werden, z.B. unter Standort „florenz via masaccio“. Es können zudem mehrere Ausdrücke eines Index mit einem weiteren Index kombiniert werden: Beispiel: Art des Werkes: „palazzo or palast“, Standort: „florenz“.

Die Treffer werden in einer knappen Strukturansicht in einem Extrafenster angezeigt (Navigator). Von dort können Vollansichten oder die Galerieansicht angesteuert werden.

Ein kritischer Aspekt bei einer Bilddatenbank ist der Schutz des Bildmaterials vor unerwünschter oder unerlaubter Vervielfältigung und Verwendung. Gerade bei kommerziell verwertbaren Bildern wie Kunstfotografien können keine Bilddateien in reproduktionsfähigen Größen ins Web gestellt werden. Deshalb entschieden wir uns für *Digilib*.

4.2 Bildpräsentation mit *Digilib*

*Digilib*³ ist eine Web-basierte Client-Server-Technologie für Bilder, die als Open-Source-Software vom *Max-Planck-Institut für Wissenschaftsgeschichte*, der *Universität Bern* und weiteren Projektpartnern entwickelt wird. *Digilib* erlaubt die Verwendung einer Vielzahl von Bildformaten, beispielsweise JPG und TIF. Im Hintergrund liegt ein Thumbnail im JPG-Format für die schnelle Galerieansicht und ein hochaufgelöstes, im Falle der bisherigen überwiegend schwarzweißen Dokumentationsfotografie durchschnittlich 20 MB große TIF-Dateien. Auf der Basis der TIF-Datei erzeugt die Software –on the fly– eine bildschirmfüllende Gesamtansicht oder etwa vom Benutzer gewünschte

³ Digilib-Homepage. URL: <http://digilib.berlios.de/>

Zooms in Teilbereiche des Bildes. So können die Betrachter der Bilder mit einer geringen Bandbreite auskommen und gleichzeitig die Zoom-Möglichkeiten bis an die Grenzen des Filmkorns nutzen. Gerade die Betrachtung der auf den konsultierbaren Fotoabzügen meist nicht oder nur kaum erkennbaren Details stellt einen großen Mehrwert der Datenbank gegenüber der konventionellen Fotothek dar. Darüber hinaus kann das Verzeichnis, das die Bilder enthält, über das Rechte-Management von *Digilib* so geschützt werden, dass ein Herunterladen der vollständigen Bilder unterbunden wird.

Digilib, das ursprünglich für digitale Bibliotheken entwickelt wurde, bietet eine Reihe von interessanten Arbeitshilfen für Kunsthistoriker:

- *Digicut* ermöglicht die Simulation eines Leuchtpultes, auf dem alle Aufnahmen zu einem Kunstwerk im Überblick betrachtet werden können.
- Die Zoom-Funktion erlaubt das Vergrößern von interessanten Bildbestandteilen, wobei der Vergrößerungsausschnitt selbst gewählt werden kann.
- Eine Annotationsfunktion erlaubt es, bis zu acht Markierungen im Bild zu setzen. Dies ist eine wichtige Arbeitshilfe für die wissenschaftliche Kommunikation bedeutet, da schnell und eindeutig auf entsprechende Markierungen im Bild Bezug genommen werden kann.
- Die Referenz-Funktion erlaubt das Versenden einer URL zu einem annotierten Bildausschnitt.

Diese Funktionen machen *Digilib* zu einem interessanten Arbeitsmittel für die kunsthistorische Arbeit online.

5 Problemfeld Digitalisierung

Zu Beginn galt es, die technische Zielsetzung (Auflösung, Farbtiefe etc.) der Digitalisierung genau zu definieren, da sich hiervon das weitere Vorgehen ableitet. Zu entscheiden war, für welche Zwecke gescannt werden sollte:

- für die volle Bildschirmanzeige,
- für Ausdruckmöglichkeit bis DIN A4 oder
- in der Absicht, die gesamte Informationsfülle der Negative im Mittelformat (6x6 cm) bis Großformat (bis 18x24 cm) zu sichern, die problemlos Vergrößerungen bis zum Plakatformat ermöglichen.

Es zeigte sich bald, daß es für uns hier keinen befriedigenden Mittelweg zwischen einer Anzeige in befriedigender Bildschirmqualität und digitaler Faksimilierung gab. Die finanzielle Ausstattung des Projekts ermöglichte – ergänzt durch Mittel des KHI und Sondermittel der MPG - eine Entscheidung zu Gunsten der zweiten Variante. Nachdem das Ziel definiert und erste Vorstellung der technischen Parameter zu dessen Erreichung in Erfahrung gebracht waren, mußte entschieden werden, ob eine interne Lösung oder die Vergabe an einen Dienstleister zu bevorzugen war.

Bereits früh bestand Einigkeit darüber, externe Hilfen oder Dienstleistungen dort zu nutzen, wo eine interne Lösung ineffektiv ist. Eine Kalkulation der damaligen Preise pro hochaufgelöstem Masterscan ergab jedoch bzgl. Kosten und Organisation langfristige Vorteile zugunsten einer internen Projektabwicklung, bei gleichzeitiger der Wahrung der Unabhängigkeit von externen Anbietern.

Auf dem anspruchsvollen und noch weitgehend unbearbeiteten Gebiet der digitalen Archivierung ließen wir uns von Prof. Rudolf Gschwind vom *Fachbereich Medienwissenschaften der Universität Basel*, beraten, einem Mitautor des Grundlagenbericht zur Digitalisierung und Sicherung fotografischer Materialien für das *Schweizer Amt für Zivilschutz* (Gschwind, Rosenthaler & Frey 2002). Bei der Entwicklung der Negative in Photoshop und der Einführung eines konsequenten digitalen Farbmanagements beriet uns Annette Keller, Fotoingenieurin der Florentiner Restaurierungswerkstätten *Opificio delle pietre dure*.

5.1 Ausstattung

Die Hardware-Ausstattung - im wesentlichen der Scanner, spezielle Monitore, Spektralometer, ein Drucker sowie ein projekteigene Server - wurden von der MPG finanziert. Unter Berücksichtigung der zu bearbeitenden Objekte (durchsichtiges Material, etwa Film oder Glas, Formate von Kleinbild bis 18 x 24 cm, hohe Dichte-Werte) und der Zielsetzung wurde ein Flextight 2848 - Scanner des Dänischen Anbieters *IMACON* angeschafft. Dieses High-End-Gerät erreicht die Qualität von Trommelscannern, erspart jedoch die umständliche Montage, und das Umkopieren der Glasnegative.

Überschlagsrechnungen gingen von einem Datenvolumens von rund 100 MB pro Master und ca. 20 MB pro Arbeitskopie aus. Die Arbeitskopien sollen dauerhaft im Hause verbleiben, während die Masterscans nur bis zum Aufbau einer gesicherten, verteilten, externen “Endlagerung” intern gespeichert

werden sollten. Für die zentral organisierte Datenhaltung wird ein Windows Server von 1,7 Terabyte mit RAID5 System verwendet. Das Backup erfolgt mit Hilfe von LTO-Bändern. Als Arbeitsplatzrechner werden 2 *Apple Power MAC G5* mit Dualprozessor sowie leistungsfähige Pentium-4-Rechner im institutsinternen Windows XP-Netzwerk eingesetzt. Entscheidend war die Wahl von farbverbindlichen *Lacie* -Monitoren und die Einführung eines Farbmanagements. Gedruckt wird auf einem *Epson*-Drucker vom Typ *Stylos Photo 2100*. Dieser ermöglicht randlose Ausdrücke bis DIN A3 und wird mit sechs Pigmenttintenpatronen sowie archivsicheren Fotopapieren betrieben . Alle Ein- und Ausgabegeräte werden regelmäßig kalibriert und ermöglichen eine konstante Farbrepräsentation.

5.2 Scanstandards, Ordnungssystem und Workflow

Wie bereits erwähnt galt es für den Aufbau des digitalen Fotoarchivs, die rein technischen Parameter für Master und Arbeitskopien festzulegen, Benennungen zu finden und daraus resultierende Ordnungs- und Speicherstrukturen einzurichten und einen neuen Workflow inklusive Qualitätssicherung auszuarbeiten. Die technischen Parameter sollen hier nur skizzenhaft beschrieben werden.

Der **Masterscan** enthält den maximalen Informationsgehalt und dient ausschließlich konservatorischen Zwecken:

- TIF-Format unkomprimiert
- 16 Bit Farbtiefe (65.536 Farben) pro Kanal (bisher nur s/w-Materialien gescannt).
- Die hohe Farbtiefe bedingt erhebliche Dateivolumen, die jedoch mit Blick auf den benötigten Spielraum für digitale Laborarbeit unbedingt erforderlich ist, andernfalls müßten starke Qualitätseinbußen akzeptiert werden.
- Punktauflösung / dots per inch (dpi):
- Die Punktauflösung richtet sich einerseits nach dem Informationsgehalt der Vorlagen (Material / Entstehungszeit) und andererseits nach dem qualitativsten zu bedienenden Ausgabemedium bzw. den dort anvisierten maximalen Abmessungen (Höhe x Breite). Folglich wurde die dpi-Anzahl für alle Formate und Materialien separat ermittelt.

Beispiele:

Vorlage	Zielmedium	Vergrößerung	Ausgabeformat	dpi für Scan
6 x 6 cm	Kunstdruck 600 dpi	x 4,5	27 x 27 cm	2.700 dpi
	Offsetdruck 300 dpi	x 9	54 x 54 cm	2.700 dpi
13 x 18 cm	Kunstdruck 600 dpi	X 2,5	32,5 x 45 cm	1.500 dpi
	Offsetdruck 300 dpi	x 5	65 x 90 cm	1.500 dpi

Unsere so genannten "**Arbeitskopien**" werden vom Masterscan ausgehend über zwei Batchverfahren und die manuelle digitale Laborarbeit in Photoshop erstellt. Angestrebt werden dabei qualitätsvolle Bilder, die für möglichst viele weitere Anwendungen (Internet, Papierausdruck, Publikation) verwendbar sind:

- TIF-Format unkomprimiert
- 8 Bit Farbtiefe
- Schmalseite per Batch auf 4000 Pixel reduziert (ca. 33cm bei 300 dpi-Druck)
- 300 dpi

Der folgende, gekürzte Dokumentations- und Digitalisierungs-**Workflow** gibt einen Überblick über den Ablauf, wobei Aspekte der Arbeitsteilung und der zeitlichen Abfolge unberücksichtigt bleiben:

- Erstellung der wissenschaftlichen Dokumentation in HiDA an Hand der „Fotopappe“ (aufgeklebter Abzug, mit Beschriftung und Stempel)
- Zusammenführung von in HiDA erfaßten Abzügen und zugehörigen Negativen aus dem Negativarchiv
- Digitalisierung der Negative und Speicherung der Masterscans in nach Tagesdatum geordneten Verzeichnissen (so genannte "Tagesordner")
- Erstellung der Arbeitskopien: Bildbearbeitung und Batchverfahren mit dem Ziel, das digitale Bild in Qualität und Ausschnitt den bestehenden Abzügen auf der Fotopappe anzupassen
- Einarbeitung der folgenden Scandaten in die bestehende Objektdatensätze in HiDA:
 - Dateiname
 - Dateiformat

Jahr und Monat der Erstellung
Name des verwandten Scanners
Urheberrechtlicher Hinweis

- In gewissen zeitlichen Abständen erfolgt eine Endkontrolle von Masterscans und Arbeitskopien, nach der die Dateien aus den "Tagesordnern" in die endgültige Verzeichnisstruktur überführt werden. Dieser Verzeichnisbaum hat zwei Hauptwurzeln für Masterscans und Arbeitskopien. Die weitere Verzweigung ist in beiden Teilen identisch und folgt der hauseigenen Nummerierung in Tausenderschritten.

5.3 Erfahrungen

Nachdem Digitalisierungsstandards, Ordnungssysteme und Workflow in wesentlichen Zügen festgelegt worden waren, erfolgte die Beschaffung der Ausstattung. Ursprünglich erschien die Geschwindigkeit des Scanners von großer Bedeutung, was sich inzwischen relativiert hat. Die Erfahrung zeigte, daß in vier Stunden nur etwa 40 Scans möglich sind, da zum einen großformatige Glasplatten nicht wie gerahmte Dias im Stapel "durchgejagt" werden können, zum anderen gerade bei Negativmaterialien zur Erzielung optimaler Ergebnisse eine individuelle Anpassung der Scaneinstellung (Weiß- und Schwarzpunkt, längere Belichtungszeit bei besonders dichtem Material, Anpassung der Gradationskurve) notwendig ist. Hinzu kommt noch etwa der gleiche Zeitaufwand für die individuelle Bildbearbeitung in Photoshop (Wahl des Bildausschnitts, Geradestellen, Ausgleich ungünstiger Lichtverhältnisse, Farb- und Kontrastabstimmung) und weitere administrativen Vorgänge, wie das Erstellen von Kontaktabzügen, Pflege von Arbeitslisten und verschiedene Kontrollvorgänge, bei denen neue aber teils auch alte Fehler zu entdecken und zu beseitigen sind.

6 Problemfeld Langzeitarchivierung

"Die Langzeitarchivierung digitaler Dokumente ist zum Problem geworden, das noch dadurch gesteigert wird, daß dieses Faktum im Allgemeinen gleichmütig hingenommen wird."; mit diesen dramatischen Worte beschreibt Hermann Leskien, Generaldirektor der Bayerischen Staatsbibliothek die aktuelle Lage. Leskien weist zu Recht darauf hin, daß die Erhaltung von digitalen Informationen über lange Zeiträume noch nicht geklärt ist und daß die Gefahr besteht, daß digitale Informationen für immer verloren gehen. Einige Autoren gehen sogar soweit, von einem heraufziehenden "digital dark age" (Kuny 1997) zu sprechen, in dem elektronisch gespeichertes Wissen

verloren geht. Dies mag etwas übertrieben erscheinen, aber es gibt tatsächlich Beispiele für Datenverluste, die zeigen, daß diese Befürchtungen nicht unbegründet sind. So gingen in den 1970er Jahren in den USA Daten der NASA über die Entwicklung des brasilianischen Regenwaldes verloren und ebenso etwa 10.000 Datensätze aus der Volkszählung von 1960 (Borghoff et al. 2003: 136). Diese Beispiele zeigen, daß es dringend notwendig ist, im Rahmen von Digitalisierungsprojekten die Frage der Langzeitarchivierung zu berücksichtigen. Unter Langzeitarchivierung wird in Anlehnung an Dobratz & Tappenbeck (2002: 257) verstanden, daß ein archiviertes digitales Objekt auch in ferner Zukunft noch verfügbar und nutzbar sein soll.

Im Gegensatz zu Büchern, die ohne Hilfsmittel gelesen werden können, benötigt man für den Zugriff auf digitale Objekte Hard- und Software als technische Hilfsmittel. Diese technischen Hilfsmittel sind für die Langzeitarchivierung ein Problem, weil sie einem rapiden Wandel unterworfen sind (Kuny 1997; Russel 1999). Dies kann dazu führen, daß schon nach relativ kurzen Zeiträumen der Zugriff auf digitale Objekte, die auf älteren Medien und in älteren Formaten gespeichert sind, erhebliche technische Schwierigkeiten bereitet. Verschiedene Forschungsprojekte legen die Schlußfolgerung nahe, daß Computer (noch) keine idealen Medien für die Langzeitarchivierung sind (Russel 1999). Denn derzeit steckt die wissenschaftliche Forschung und Diskussion noch in den Anfängen und die Frage nach den Möglichkeiten und Bedingungen einer zuverlässigen Langzeitarchivierung ist heute noch weitgehend unbeantwortet (Dobratz & Tappenbeck 2002: 257).

6.1 Methoden der Langzeitarchivierung

Für die Langzeitarchivierung digitaler Objekte sind vier Methoden in der Diskussion:

- Erstellung von Hardcopies: es werden Ausdrücke von digitalen Informationen erstellt und in Papierform archiviert. Der Vorteil ist, daß man über umfangreiche Erfahrung in der Archivierung von Papier verfügt. Der Nachteil ist, daß mit dem Ausdruck digitaler Objekte nicht nur ihre Authentizität verloren geht, sondern auch ein mehr oder weniger großer Teil ihrer Funktionalität (z. B. bei Hypertexten oder hypermedialen Anwendungen). Deshalb wird dieser Ansatz von Kritikern als nicht akzeptabel bezeichnet.
- Erhaltung von Hard- und Software: dieser Ansatz geht davon aus, daß von allen verfügbaren Computermodellen ein voll funktionsfähiges Exemplar bereit gehalten wird, damit künftige Generationen bei Bedarf

digitale Objekte in ihrer ursprünglichen Wiedergabeumgebung betrachten können. Der Vorteil wäre eine nicht zu übertreffende Authentizität der Archivierung, weil die Originalwiedergabeumgebung verwendet wird. Ein möglicher Nachteil könnte sein, daß potentielle Benutzer in 100 Jahren nicht mehr mit der Originalwiedergabeumgebung vertraut sind und für den Zugriff erst geschult werden müßten. Ein weit größerer Nachteil ist der Aufwand, der betrieben werden muß, um die Original-Hardware dauerhaft funktionsfähig zu halten. Kritiker bestreiten, daß dies in der Praxis möglich und finanzierbar ist.

- Datenmigration: bei der Migration wird versucht, die Daten ständig auf dem aktuellen Stand der Darstellungsform zu halten. Dazu werden sie in bestimmten zeitlichen Abständen auf die nächste Computergeneration überspielt.

Der Vorteil ist, daß die digitalen Objekte immer in einer aktuellen Form vorliegen und bearbeitbar sind. Der Nachteil ist, daß es bei der Migration zu Veränderungen der digitalen Objekte kommen kann und daß ursprüngliche Informationen verloren gehen können. Weiterhin müssen für jeden Migrationsschritt alle Objekte bearbeitet werden, so daß die Migration im Laufe der Zeit immer aufwendiger wird. Hierin sehen Kritiker ein Realisierungs- und Finanzierungsproblem.

- Emulation: bei diesem Ansatz werden die digitalen Objekte in ihrer Originalform gespeichert, aber auf die Erhaltung der Original-Hardware wird verzichtet. Statt dessen wird die Funktionsweise der Hardware auf einem künftigen Computer nachgebildet (emuliert). Es werden also nicht die digitalen Objekte migriert, sondern die Wiedergabeumgebungen, weil nicht praktikabel erscheint, die Original-Hardware dauerhaft funktionsfähig zu halten. Der Vorteil der Emulation ist, daß das Originalobjekt unverändert verfügbar bleibt und der Aufwand für die Emulation - im Gegensatz zur Migration - unabhängig von der Anzahl der Objekte ist. Der Nachteil der Emulation ist, daß künftige Computer, die aktuellen Computer simulieren sollen, erheblich leistungstärker sein müssen als die aktuellen, weil durch die Emulation Leistung verloren geht. Kritiker bestreiten die praktische Umsetzbarkeit der Emulation wegen technischer und urheberrechtlicher Probleme.

Für weiterführende Informationen bezüglich dieser Methoden wird auf Borghoff et al. (2003) verwiesen.

Nach dem derzeitigen Stand der Kunst werden für die Archivierung digitaler Objekte häufig so genannte kombinierte oder Hybridlösungen befürwortet, bei denen die Informationen auf zwei Arten archiviert werden (Russel 1999; Borghoff et al. 2003: 144):

- für die Kurzzeitarchivierung werden die Informationen in digitaler Form aufbewahrt, um die Vorteile des elektronischen Zugriffs und der digitalen Verbreitung und Verarbeitung zu haben und
- für die Langzeitarchivierung werden die Informationen auf Mikrofilm gespeichert.

Das KHI hat sich hingegen für eine rein digitale Langzeitarchivierung auf Basis der Datenmigration entschieden. Einerseits sind die analogen Originale bei entsprechender fachgerechter Lagerung bereits relativ beständig. Die gängigen Mikrofilmformate würden andererseits zu einer wesentlichen Reduzierung des Informationsgehaltes führen. Die Kritiker der ausschließlich digitalen Sicherung argumentieren häufig mit der Größe der entstehenden Datenvolumina, ferner Schwierigkeiten bei der Verwaltung und den letztlich hohen Kosten. Unsere Kalkulationen hinsichtlich des Datenvolumens gehen zwar in den Terabyte-Bereich, sind aber angesichts der rasanten Entwicklung auf dem Gebiet der Speichertechnologien zunehmend als unproblematisch einzuschätzen. Auch die Verwaltung und Ordnung der digitalen Daten sehen wir zumindest in unserem Falle mit keinen besonderen Schwierigkeiten verbunden. Allerdings gehört die technische Gewährleistung einer derartigen Langzeitarchivierung nicht zu den Kernaufgaben eines kunsthistorischen Forschungsinstitutes. Hier wurde nach einer externen Lösung gesucht.

6.2 Konzept der digitalen Datenhaltung

Für die langfristige Datenhaltung und –archivierung ist die Zusammenarbeit mit zwei großen externen Rechenzentren begonnen worden, nämlich mit der *Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG)* in Göttingen und dem *MPI für Plasmaphysik* in Garching bei München. Durch die Kooperation mit diesen Partnern soll die ständige Verfügbarkeit der Daten, ihre sichere Aufbewahrung sowie ihre zeitgerechte Migration sichergestellt werden.

Der Datentransfer von Italien nach Deutschland erfolgt über das italienische bzw. deutsche Wissenschaftsnetz. Zur Zeit können hierbei 512 Kbit / sec ausgenutzt werden. Dies entspricht etwa 5 GB in 24 Stunden. Für den Abgleich der Datenstände und der Transfer wurden *rsync*-Prozesse eingerichtet. In Göttingen befinden sich eine Online-Version und zwei Band-Backups an räumlich getrennten Stellen. Ein *AFS-Client (Andrew File System – Client)* ermöglicht jeder Zeit eine Einsichtnahme und Kontrolle und gegebenenfalls Rückübertragung von Daten am Garchinger Standort. Dort existieren zwei getrennte Versionen in großen Bandrobotersystemen. Neben

dem Transfer über Leitungen wurden auch größere Datenmengen auf USB Platten gespeichert und über Kurierdienste versandt, der normale Postweg ist nicht zu empfehlen. Checksummen dienen als Kontrollinstrument für eine fehlerfreie Übermittlung der Daten. Die Verzeichnisstruktur ist systembedingt offen angelegt, da nicht systematisch bestimmte Nummernbereiche gescannt wurden, sondern die Auswahl inhaltlichen Kriterien folgte. So können jederzeit weitere Dateien in schon bestehende Verzeichnisse abgelegt werden oder auch weitere Verzeichnisse entstehen

6.3 Erfahrungen und Ansätze in der Langzeitarchivierung

Die Problematik der Langzeitarchivierung dringt erst langsam ins allgemeine Bewußtsein. Wenige Einrichtungen insbesondere im geisteswissenschaftlichen Bereich haben bisher direkt mit dem Ziel einer Langzeitsicherung digitalisiert. In jüngster Zeit taucht das Thema auf den Vortragslisten von Tagungen auf und der Aufbau des zentralen Kompetenznetzwerkes *NESTOR*⁴ wurde in Angriff genommen. Auch innerhalb der Max-Planck-Gesellschaft wird hier zunehmend Handlungsbedarf für die Bereitstellung zentraler Speichermöglichkeiten und Migrationskonzepte gesehen. Die GWDG ist deshalb dabei, in Kooperation mit verschiedenen Max-Planck-Instituten den Bedarf zu analysieren und ein Konzept für die Max-Planck-Gesellschaft zu entwickeln (Ullrich 2004). Im Mittelpunkt steht dabei die „Bitstream Preservation“ als zuverlässige langfristige Speicherung der Datenobjekte. Darunter wird die langfristige Speicherung der Datenobjekte verstanden.

Zwei Schlüsselthemen im Zusammenhang mit der Herausforderung Langzeitarchivierung sind die technische Gewährleistung und die Frage der Verantwortung. Sicherungszeiträume von dreißig, fünfzig oder auch mehr Jahren werden häufig genannt. Technisch überschaubar ist jedoch allenfalls die nächste Rechner- und Speichergeneration, womit ein Zeitraum von eventuell drei bis fünf Jahren gemeint ist. Aussagen über längere zukünftige Zeiträume beruhen auf der Extrapolierung vergangener Entwicklungen und Erfahrungen. Mit dem Anwachsen der zu bewahrenden Daten stellt sich auch die Frage nach dem Träger der Verantwortung. Das KHI erzeugt zwar Daten von kulturhistorischer und wissenschaftlicher Bedeutung wünscht sich für die verantwortungsvolle Sicherung jedoch Unterstützung von erfahrenen Rechenzentren. Diese sehen ihrerseits die Frage der Verantwortung als nicht gelöst an und erhoffen sich eine Klärung auf nationaler und politischer Ebene.

⁴ Kompetenznetzwerk Langzeitarchivierung URL: <http://www.langzeitarchivierung.de/>

7 Zusammenfassung

Die Erfahrung aus dem Digitalisierungsprojekt hat gezeigt, dass der Aufwand für ein solches Projekt nicht unterschätzt werden darf (vgl. Pfenniger 2001: 10). Gerade die Digitalisierung der Archivbestände erfordert umfangreiche Vor-, Begleit- und Nacharbeiten, die den bisherigen Arbeitsablauf in Frage stellen und organisatorische Änderungen notwendig machen können. Bei der Gestaltung der Suche in der Bilddatenbank stellte sich die hierarchische Struktur der Bildmetadaten aus HiDA als eine besondere Herausforderung dar. Durch verschiedene Indizes konnte aber ein umfassender Zugriff auf die Bilddaten ermöglicht werden. Mit Digilib bietet die Bilddatenbank ein nützliches Werkzeug für die wissenschaftliche Arbeit online. Die Frage der Langzeitarchivierung konnte noch nicht abschließend beantwortet werden. Mit einer parallelen Datenhaltung in zwei Rechenzentren und der Teilnahme an den Langzeitarchivierungsplänen der MPG wurde ein Weg beschritten, der zukunftsfähig erscheint.

8 Literatur

- Borghoff, Uwe M./Rödig, Peter/Scheffczyk, Jan/Schmitz, Lothar (2003): Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente. Heidelberg: dpunkt Verlag.
- Dobratz, Susanne/Tappenbeck, Inka (2002): Thesen zur Zukunft der digitalen Langzeitarchivierung in Deutschland. In: Bibliothek 26 (3) 2002: 257-261.
- Bove, Jens/Heusinger, Lutz/Kailus, Angela (2001): Marburger Informations-, Dokumentations- und Administrations-System (MIDAS). (Literatur und Archiv Bd. 4) München: Saur
- Gschwind, Rudolf/Rosenthaler, Lucas/Frey, Franziska (2002): Digitale Archivierung fotografischer Sammlungen, ein Grundlagenbericht. Internet, URL <http://www.abmt.unibas.ch/downloads/KGS-Grundlagenbericht.pdf>. Zugriff: 30.07.04.
- Kuny, Terry (1997): A Digital Dark Ages? Challenges in the Presevation of Electronic Information. In: Proceedings of the 63rd IFLA Council and General Conference. Internet, URL <http://www.ifla.org/IV/ifla63/63kuny1.pdf>. Version: 08/27/1997. Zugriff: 30.07.04.
- Pfenniger, Kathryn (2001): Bildarchiv digital. (Rundbrief Fotografie; Sonderheft 7) Esslingen: Museumsverband Baden-Württemberg.
- Russell, Kelly (1999): Digital Preservation: Ensuring Access to Digital Materials Into the Future. Internet, URL <http://www.leeds.ac.uk/cedars/Chapter.htm>. Version: June 1999. Zugriff: 30.07.04.
- Ullrich, Dagmar (2004): Digitale Langzeitarchivierung. In: GWDG- Nachrichten für die Benutzer des Rechenzentrums, 27. Jg. 7/2004. Internet, URL <http://www.gwdg.de/forschung/publikationen/gwdg-nr/GN0407/gn0407.pdf>. Version: 07/2004. Zugriff: 30.07.04.