

EVA *Europe* *'99 Berlin*

Elektronische Bildverarbeitung & Kunst, Kultur, Historie

Die 6. EVA Berlin ist 1999 die europäische Hauptveranstaltung
der weltweit etablierten EVA-Serie

Electronic Imaging & the Visual Arts

9. - 12. November 1999

$\frac{S}{P} \mid \frac{M}{K}$

bei den Staatlichen Museen zu Berlin-
Preußischer Kulturbesitz
am Berliner Kulturforum



Gesellschaft zur Förderung
angewandter Informatik e.V.

VASARI UK

Tutorial 2

Digitalisierungsstrategien und Aufbau digitaler Archive

Projektmanagement - Strategien der Digitalisierung im Bereich Erfassung, Forschung und Vermittlung Alfred Weidinger (Graphische Sammlung Albertina, Wien)	T 2.1
Informationsbedürfnis am 'cultural heritage' / Digitalisierung von Museumsbeständen / Berücksichtigung von Informationsanforderungen / Einsatz gewonnener Daten / Datenzugriff und -transfer	
Vom Bildarchiv zum digitalen Bildarchiv/Bildindex Carsten Reimann, Jens Bove (Bildarchiv Foto Marburg)	T 2.2
Transfer unter Beibehaltung der Ordnungssysteme / Digitalisierung: vorlagenorientiert / Speichertechniken / Software-Einsatz: Bildverarbeitung / Migration: Techniken, Zyklen / Verfahren der Zusammenarbeit / Ausgabeprodukte für Werbung und Verkauf	

PROJEKTMANAGEMENT
für die Datenaufnahme und Digitalisierung von Kunstwerken
am Beispiel der
Graphischen Sammlung Albertina, Wien

Sammlungsbestände:

Graphische Sammlung Albertina:	40.000 Zeichnungen und Aquarelle, 1 Mio Druckgraphiken
Architektursammlung Albertina:	25.000 Zeichnungen und Pläne
Fotosammlung Albertina:	30.000 Fotografien, 500 hist. Fotoapparate

Projektbeginn: 10/1998

Inhalt des Vortrags:

1. Einleitung
2. Analyse 1998
3. Definierung der Ziele, Zeithorizont
4. Schaffung der techn. und räuml. Voraussetzungen, Personalrekrutierung
5. Projektbeginn
6. Probleme in der Anfangsphase
7. Ablöse der bestehenden Datenbank
8. Erfahrungen mit der neuen Datenbank
9. Projekterweiterung
10. Vorbereitungen zu Erfüllung des ersten Zwischenziels
11. Investitionen, Kosten
12. Abschlußbemerkung

Referent: Alfred Weidinger Kurator der Albertina
Leiter der Stabsstelle für Infrastruktur- und Baumanagement
Beauftragter des Bundesministeriums für Unterricht und
kulturelle Angelegenheiten

Digitalisierung eines Bildbestandes

Vor dem Aufbau eines digitalen Bildarchivs sind viele Entscheidungen zu treffen und viele Fragen zu beantworten, z.B.:

Welche Bestände eignen sich für die Digitalisierung?

Nachdem geklärt ist, welche - möglichst geschlossenen - Bestände interessant für die Öffentlichkeit sein könnten, muß analysiert werden, ob geeignete Bildvorlagen sowie erschließendes Datenmaterial vorhanden sind und wer die Rechte daran hat. Aber vor allem muß entschieden werden, zu welchem Zweck und in welchem Medium die Bilder angeboten werden sollen. Daran anknüpfend stellt sich die Frage, auf welche Weise die Vorlagen digitalisiert werden sollen, welche Speichermedien zur Verfügung stehen bzw. finanzierbar sind und welche Software eingesetzt werden soll. Dabei muß berücksichtigt werden, ob die Migrationsfähigkeit der Informationen sichergestellt ist und nicht zuletzt muß überprüft werden, welche Chancen der Zusammenarbeit wahrgenommen werden können.

Dieses Papier und - detaillierter - das Tutorial verstehen sich als Anregung, die richtigen Fragen zu stellen und verschiedene Vorgehensweisen vorzustellen. Im einzelnen:

Sind (Teil-) Bestände bereits erschlossen, welche Art von Datenmaterial existiert?

Weit teurer und aufwendiger als die Erzeugung und Bereitstellung digitaler Bilder ist ihre Erschließung. Daher sollte zuerst geprüft werden, welche Art von Informationsmaterial jeweils verfügbar ist. Nicht in allen Museen liegen gedruckte Bestandsverzeichnisse vor; oft stehen nur Auswahlkataloge zur Verfügung. Insbesondere Deposita sind in vielen Häusern nur durch - vielfach handschriftliche - Inventarlisten oder Karteikarten erschlossen. EDV-gestützte Katalogisierung muß eher als Ausnahme betrachtet werden. Auch große Teile von Bibliotheks- und Archivbeständen sind nur über vielfach heterogene Karteisysteme zugänglich. Nur selten liegen vor dem Aufbau eines digitalen Bildarchivs bereits flächendeckend maschinenlesbare Daten vor.

Entscheidend für die Finanzierung und Realisierung eines Projekts ist also die Frage, inwieweit vorhandene Informationen ökonomisch in eine Datenbank transferiert werden können. Grundsätzlich bieten sich zwei Möglichkeiten, schriftliche Vorlagen in maschinenlesbare Daten zu konvertieren: maschinengestützte Retrokonversionsverfahren oder personalintensive Neukatalogisierung.

Die Entscheidung für ein bestimmtes Retrokonversionsverfahren ist abhängig von den jeweils vorhandenen finanziellen und personellen Ressourcen, der spezifischen lokalen Situation, der Menge der zu konvertierenden Daten, Art und Umfang der nutzbaren Fremddaten (z.B. Normdateien wie ICONCLASS) sowie den mit dem Ergebnis der Konversion verbundenen Zielen und Ansprüchen.

- **OCR**
Die Erfahrungen des Bildarchivs mit automatischer Texterkennung (OCR) sind bisher negativ. Hohe Fehlerquoten erfordern ein enormes Maß an Nachbearbeitung, insbesondere bei sehr heterogenen Vorlagen und einer hohen Zahl von Textauszeichnungen. Eine Alternative ist die Verbindung unbearbeiteter OCR-Ergebnisse mit gesicherten Metadaten (s.u.) unter Inkaufnahme fehlerbehafteter Datensätze.
- **Image-Digitalisierung**
Ein relativ preiswerter Weg zur Integration alter Informationsbestände in eine Datenbank ist die Sicherheitsverfilmung und anschließende Digitalisierung von Textvorlagen als Bilddateien in Verbindung mit geeigneten Metadaten (s.u.).
- **Schreibbüro**
Abschreiben von kompletten Vorlagen durch Schreibbüros, evtl. in Billig-Lohn-Ländern, oder Abschreiben relevanter, evtl. vorstrukturierter Teile und anschließende Nachbearbeitung zur gezielten Erschließung von Text- und Bildinformationen.
- **bereits vorhandene Datenbank**
Elektronisch vorliegende Daten müssen auf ihre Kompatibilität mit der einzusetzenden Software des digitalen Bildarchivs geprüft und ggf. entsprechend der gestellten Anforderungen redaktionell nachbearbeitet werden.

Wie produziere ich Metadaten?

Die Produktion von Metadaten ist der wohl schwierigste und weitaus aufwendigste Teil beim Aufbau eines digitalen Bildarchivs. Einfache Rezepte sucht man hier vergeblich, da die zu wählende Strategie in erster Linie von den bereits vorhandenen Materialien ausgehen muß; die Lösungen sind im Detail immer sehr individuell. In jedem Fall empfiehlt sich jedoch die Übernahme bestehender Ordnungssysteme. Anhand von Projekten, an denen im Bildarchiv Foto Marburg zur Zeit gearbeitet wird, können drei mögliche Produktionswege von Metadaten beispielhaft erläutert werden.

- Das "Bildarchiv zur Kunst und Architektur" (BKA) wird in drei Jahren 1,5 Millionen kunstgeschichtlicher Fotografien bieten und zumindest einen Teil der reproduzierten Kunst- und Bauwerke unter verschiedenen Gesichtspunkten erschließen. Der gegenwärtige Prototyp 3 enthält 377.000 Fotografien aus Orten von Aach bis Esslingen. Der Aufbau des digitalen Bildindex orientiert sich an seinen Vorlagen, den im "Marburger Index" publizierten und jetzt digitalisierten Microfiches. Die Präsentation im Internet übernimmt das Ordnungssystem des „Marburger Index“. Es führt von Leitkarten, durchschnittlich eine Leitkarte für zehn Microfiches, über differenzierte Titelleisten der Fiches zu den fotografischen Reproduktionen mit bestandsgliedernden Zwischenüberschriften und mehr oder weniger ausführlichen Beschriftungen. Obligatorische Mindestbestandteile der Beschriftungen sind eine Ortsangabe und genaue Lokalisierung des Objekts, eine genaue Sachbezeichnung und die Nennung von Eigentümer, Negativnummer und Aufnahmedatum der Fotografie. Vgl. <http://bildindex.de>
- Die Handschriftendatenbank enthält 27.000 Handschriftennachweise, die aus den bereits in Datenbankform vorliegenden Registereinträgen des "Gesamtindex mittelalterlicher Handschriften" erzeugt und mit digitalen Reproduktionen der Handschriftenbeschreibungen aus 150 gedruckten Katalogen verknüpft worden sind. Da die Registereinträge dieser Kataloge in ihrer digitalen Ausgangsform, ihrer ursprünglichen Funktion entsprechend, weder kategorisiert noch einheitlich sind, mußte diese redaktionelle Arbeit erst geleistet werden. Mittelfristig sollen die Daten außerdem mit digitalen Reproduktionen der bereits auf Microfilm vorliegenden Handschriften verbunden werden. Vgl.: <http://www.fotomr.uni-marburg.de/Handschriften-Forum.htm>
- In einem weiteren Schritt sollen auch ältere Kataloge und Inventare in die Handschriftendatenbank integriert werden. Im Gegensatz zu den oben genannten Vorlagen liegen hier keine maschinenlesbaren Daten vor, so daß die Metadaten zur Erschließung der Text-Images gänzlich neu produziert werden müssen. Geplant ist, die Registereinträge - soweit vorhanden - abschreiben zu lassen und nach dem obigen Muster zu kategorisieren und zu vereinheitlichen. Liegen keine Register vor, was der Regelfall ist, werden die Überschriften und Schlagzeilen der einzelnen Katalogisate vorstrukturiert, abgeschrieben, auf Datenbankbasis vereinheitlicht und schließlich mit den Images der Textvorlagen verknüpft.

Wer ist Eigentümer der Bildrechte?

Vor einer Digitalisierung und Veröffentlichung eigener Bildbestände muß geprüft werden, inwieweit die Sammlung oder einzelne Teile urheberrechtlich geschützt sind. Dabei sollte beachtet werden, daß einige Kollegen die Meinung vertreten, daß fotografische Reproduktionen zweidimensionaler Vorlagen grundsätzlich keinem Urheberrechtsschutz unterliegen, soweit die Vorlagen selbst gemeinfrei sind; in der Regel 70 Jahre nach dem Tod des Urhebers. Ob Museen, Bibliotheken oder Archive über dieses selbstverständliche Urheberrecht hinaus ein Verfügungs- oder Veröffentlichungsrecht bzw. "Copyright" geltend machen dürfen, sofern es sich um öffentliche Institutionen handelt, ist ebenfalls nicht endgültig geklärt. Es ist keineswegs unumstritten - insbesondere im Bereich Internet -, ob öffentliche Einrichtungen die Befugnis haben, bildliche Wiedergaben der von ihnen verwalteten Bestände kommerziell nach dem Muster privater Eigentümer mit Lizenzgebühren zu belegen, die über die Herstellungskosten für Reproduktionsmaterial hinausgehen.

Zu den gegensätzlichen Positionen vgl.:

<http://www.fotoinfo.de/fotorecht>

<http://www.uni-koblenz.de/~graf/kultjur.htm>

Wie digitalisiere ich meine Bildbestände?

Nach der Analyse der vorliegenden Daten folgt der eigentliche Schritt der Digitalisierung. Art und Umfang richten sich sowohl nach vorlageorientierten wie nach ausgabeorientierten Kriterien.

- **vorlagenorientierte** Aspekte:

Zunächst muß gesichtet werden, welches Material vorliegt. In welchem Umfang kann man auf bereits verfilmte Bestände zugreifen und in welchem Umfang muß neuverfilmt werden?

Bei der **Neuverfilmung** ist zu klären, welche Verfahren sich anbieten. Farbfotos sind oft gerade für den Printbereich erheblich attraktiver als Schwarzweissfotos. allerdings lassen sich nur ganz bestimmte Farbfilme langfristig archivieren. Weitere Möglichkeiten wären Microverfilmung oder das direkte Digitalisieren mit einer Digitalkamera ohne Umweg über die konventionelle Fotografie. Welches Verfahren bietet hier welche Vorteile?

In kleinem Rahmen ist es durchaus Möglich, Scans selber im Haus mit relativ einfachen Mitteln durchzuführen. Ab einer gewissen Menge ist es aber ratsamer, diesen Bereich von externen Firmen erledigen zu lassen. Dabei ist vor allem auch die Art der Vorlage ausschlaggebend. Welche Techniken und Dienstleister gibt es in diesen Bereichen, worauf ist zu achten? Erfahrungen und Probleme?

- **ausgabeorientierte** Aspekte:

Für die Digitalisierung ist ganz entscheidend, für welche Endanwendung der Scan benötigt wird. Für eine digitale Veröffentlichung z.B. auf CD oder im Internet braucht man natürlich qualitativ viel weniger anspruchsvolle Scans als für die Ausgabe im Print-Bereich. Dementsprechend muß man auf unterschiedliche Dateiformate und Speichermedien zurückgreifen, die wiederum das Datenaufkommen bestimmen. Für welche Anwendung benötigt man welche Qualität und welche Bildformate?

Zudem ist zu überlegen, ob ein einmaliges, die fotografischen Informationen ausschöpfendes Scannen geplant und möglich ist oder ob rein projektorientiert (z.B. für eine CD-Rom) für eine bestimmte Anwendung Material in ganz anderer Qualität einmalig benötigt wird.

Wie speichere ich meine Bilder?

Die angelieferten Scans müssen i.d.R. vom Transfermedium (CD) herunter auf einen Server/Arbeitsplatz kopiert werden, um weiterverarbeitet und/oder angeboten werden zu können. Erhebliche Datenmengen können anfallen. Wohin also mit dem digitalen Bild?

- **Bildformate:** Eine Grundsatzentscheidung stellt oft schon das Bildformat, in dem die Digitalisate gespeichert werden, dar. Welche verschiedenen Formate gibt es, was sind die Unterschiede und was bietet sich für welchen Zweck an? Ist eine Datenkompression vertretbar oder sollte verlustfrei gespeichert werden? Wie kompatibel sind solche Bildformate, gerade im Bezug auf eine längerfristige Archivierung?
- **Speichermedien:** Welche physikalischen Medien bzw. Datenträger für Transfer, Speicherung und die Archivierung gibt es? Scans werden i.d.R. auf CD's angeliefert und danach auf Festplatte kopiert. Wie sieht es dabei aus mit Haltbarkeit und Formaten, z.B. bei einem Wechsel des Betriebssystems? Ab einem bestimmten Datenvolumen sind Festplatten als Speichermedium nicht mehr wirtschaftlich. Hier treten Tape-Roboter oder DVD-Roms/Rams an die Stelle. Welche Techniken gibt es in diesem Bereich und mit welchen Kosten muß man rechnen?
- **Hardware:** Wie sollte ein Server konzipiert sein, um die Informationen den jeweiligen Vorgaben entsprechend zu speichern und anzubieten? Am Beispiel des langsam wachsenden Servers des Bildarchivs Foto Marburg lassen sich die unterschiedlichen Konzepte und Ausbaustufen gut nachvollziehen. Aus einem einfachen Netzwerkserver wurde zunächst ein Internetserver und demnächst ein Storage-Server mit rund 10 Terrabyte online.

Welche Software kann ich für welchen Zweck einsetzen?

- **Bildverarbeitung:** Oft bedürfen eingescannte Bilder einer Nachbearbeitung. Ränder bzw. Bilderrahmen müssen entfernt, Bildgrößen meist in verschiedene Auflösungen reduziert, Farbstiche und Tonwerte korrigiert werden. Zusätzliche Filterfunktionen der Bildverarbeitungssoftware, wie 'Scharfzeichnen' oder 'Staub & Kratzer entfernen' bieten Möglichkeiten, die Bilddarstellung auf dem Monitor zu optimieren. In welchem, vor allem quantitativen Rahmen, kann man Bilder nachbearbeiten und was für Gefahren bestehen dabei? Welche Softwareprodukte bieten sich an? Braucht man spezielle (z.B. farbkalibrierte) Arbeitsplätze?
- **Workflow:** Bei großen Datenmengen ist ein automatisierter Transfer notwendig: Wie kommen Bilder von der CD in der richtigen Auflösung auf die richtige Platte und in das entsprechende Verzeichnis? Welche Möglichkeiten bieten Batch-Programme, wann sollten individuelle Lösungen programmiert werden?
- **Ausgabesoftware:** Welche Produkte oder Leistungen sollen letztendlich angeboten werden? Kann man dabei auf bestehende Softwarelösungen zurückgreifen oder müssen hier Firmen mit der Programmierung beauftragt werden?
- **CD-Rom-Produktion:** Hier gibt es zahlreiche Lösungen am Markt, die das Erstellen einer CD mit allen Möglichkeiten wie Menüs, Requestern etc. via grafischer Oberfläche scheinbar spielend leicht machen, wie z.B. Macromedia Director. Schwierig dabei ist aber vor allem das Einbinden von Datenbankabfragen. Dabei ist der Aufwand für eine CD-Rom nicht zu unterschätzen; viel KnowHow und Manpower sind nötig. Alternativ können solche Aufgaben von Firmen übernommen werden. Mit welchem Aufwand ist zu rechnen?

Im Vergleich: Augustinermuseums-CD per Shareware (lowcost) // 'Dürers Dinge' (Macromedia) oder 'Die Kunst der 60er' (HTML) als MidPrice-Variante // Diskus-CD-Reihe mit speziell entwickelter Softwarelösung (highend).

- **Intranet-Lösung:**

In der Regel sind Intranetlösungen direkt an Verwaltungsfunktionen der jeweiligen benutzten Datenbank geknüpft. Bearbeiter haben so jederzeit die Möglichkeit, auf Text und Bildinformationen zuzugreifen. Ist die kombinierte Funktion softwaretechnisch nicht vorgesehen, kann eventuell eine Internet-Lösung für das Intranet genutzt werden.

- **Internet-Angebot:**

Via Internet kann ein Bestand weltweit vermittelt werden. Die dazu nötige Programmiersprache HTML zum Bau von Internetseiten ist zwar sehr leicht zu lernen, darüber hinaus gehende Funktionen wie z.B. die Kommunikation mit Datenbanken oder die Anzeige von großen Bildmengen sind aber oft nur mittels programmierter Speziallösung möglich.

Im Vergleich: Internet-Handschriften-Katalog (<http://www.fotomr.uni-marburg.de/hs/hs-online.htm>): Angebot eines Bildbestandes als reine HTML-Lösung und

Bildarchiv zur Kunst und Architektur (<http://bildindex.de/intro.htm>): automatisch generierte Internetseiten mit programmierter Datenbankanbindung.

Unter welchen Voraussetzungen sind meine Daten und Bilder migrationsfähig?

Bild- und Textdaten sind flüchtig, da sowohl Soft- als auch Hardwaresysteme ständigen Veränderungen unterliegen. Um sicherzustellen, daß digitale Bilder und Daten nach einem Wechsel, wie er in der Vergangenheit etwa von DOS zu Windows, von 16bit zu 32bit Systemen stattgefunden hat, weiterhin - und ohne Qualitätsverlust - lesbar sind, müssen entsprechende Maßnahmen beim Aufbau eines Bildarchivs berücksichtigt werden.

Eine Strategie ist die strikte Einhaltung von Industriestandards, deren Aufwärtskompatibilität als sichergestellt gilt. Für Bilddateien gelte dies vor allem für das speicherplatzintensive TIFF-Format. Eine Alternative, die das Vorhalten großer Datenmengen unter Umständen ersetzen kann, ist die grundsätzliche Anfertigung konventioneller Sicherheitsverfilmungen auf lagerfähigem Microfilm. Dieser steht bei Bedarf langfristig als Basis für eine einfache und relativ preisgünstige Neudigitalisierung zur Verfügung.

Einziger relevanter Standard für Textinformation ist steuerzeichenfreier ASCII-Code, der in jedes Textverarbeitungs- und Datenbanksystem importierbar ist. Bei der Wahl geeigneter Software für ein Bildarchiv ist also auch zu prüfen, ob alle Daten ohne Informationsverlust in reinen ASCII-Code exportierbar sind.

Eine andere Möglichkeit bietet die Textspeicherung in Form von Bilddateien im TIFF-Format. Diese Methode bietet sich in erster Linie für abgeschlossene, unveränderliche Dokumente wie Rechnungen, Briefe etc. sowie für Retrokonversionsvorhaben an. Für Datenbankinhalte, die ständiger redaktioneller Bearbeitung unterliegen, ist diese Lösung nicht geeignet, allerdings könnten ältere Zustände von Datensätzen im Rahmen einer Versionierung von Datenbankinhalten auf diese Weise dokumentiert werden. Unabdingbar ist bei diesem Verfahren jedoch immer die Erschließung der Text-Images durch geeignete Metadaten.

Brauche ich Partner für die Realisierung meines Projekts?

Vielfach hat sich die Erkenntnis durchgesetzt, daß Fortschritte in der kunsthistorischen und bibliothekarischen Dokumentation und Materialerschließung heute nur durch interinstitutionelle Zusammenarbeit erzielbar sind. Ziel sollte sein, überinstitutionell für alle Interessierten - möglichst kostenfrei - nutzbare leistungsfähige und flexible neue Informationsinstrumente zu etablieren, deren Wert aus ihrer wechselseitigen Verknüpfbarkeit resultiert.

Als Beispiel der eigenen Initiativen sei die Verbunddatenbank auf dem Server des Bildarchivs Foto Marburg genannt. In ihr sind aus der "Datenbank des Bildarchivs zur Kunst und Architektur", der "Handschriftendatenbank", der "AKL-Datenbank" und der "Forschungsdatenbank" insgesamt 413.000 Dokumente zusammengeführt, die in einer Vielzahl von Archiven, Bibliotheken, Denkmalämtern und Museen in Orientierung an dem Regelwerk des Marburger Informations-, Dokumentations- und Administrations-Systems (MIDAS) angelegt worden sind und auf 130.000 Normdokumenten und einem einheitlichen hierarchischen Thesaurus beruhen.

Der Prototyp der Verbunddatenbank dient der Verbesserung von Verfahren für den Aufbau und die wirksame Erschließung umfangreicher kulturwissenschaftlicher Dokumentationsbestände unterschiedlichen Inhalts im Rahmen der Errichtung der verteilten digitalen Forschungsbibliothek der Deutschen Forschungsgemeinschaft.

Vgl. <http://www.fotomr.uni-marburg.de/vb.htm>

